3 1761 09702660 3

Columbia University

Contributions to Education

Teachers College Series

Psych
S694

w
STORAGE

hilos.
sychol.
S.
Psych
S694

PEDAGOGICAL PROGNOSIS

PREDICTING THE SUCCESS OF PROSPECTIVE TEACHERS

By

GROVER THOMAS SOMERS, Ph.D.

Teachers College, Columbia University
Contributions to Education, No. 140

192960
11.12.24

Published by
Teachers College, Columbia University
New York City
1923

Copyright, 1923
By
GROVER THOMAS SOMERS

ACKNOWLEDGMENT

This investigation from the beginning has been a coöperative project in a very real sense,—the undertaking being made possible only by the participation of many persons,—and credit for assistance rendered is widespread. Space does not permit of the enumeration individually of all those who contributed to its success. The author wishes especially to acknowledge his indebtedness to President J. L. Jarman and the faculty of the State Normal School for Women, Farmville, Va., to Dr. W. F. Tidyman and his training school staff, to the supervisory force in the field, and to the "girls" who served as subjects for the study,—without whose coöperation the data could not have been assembled nor the work accomplished.

The writer also desires to express his appreciation to his advisory committee under whose guidance the report has been prepared. For whatever of especial value the treatment by partial and multiple correlation and the regression equation may possess, credit is due Professor H. A. Ruger for introducing the author to the method and for piloting him through its difficulties; to Professor A. I. Gates for stimulating and constructive suggestions in the treatment of the data; and to Professor W. C. Bagley for the correct emphasis placed on character traits and qualities related to success in school and outside, and for his painstaking examination of the manuscript and helpful criticism throughout.

And finally the writer wishes to acknowledge the unmeasured debt to his Mother, without whose counsel and encouragement the study would not have been undertaken and could not have been carried through.

G. T. S.

Digitized by the Internet Archive
in 2014

https://archive.org/details/pedagogicalprogn00some

CONTENTS

PAGE

I INTRODUCTION 1

II. THE STUDY AS MADE 7
  Traits to be Measured 8
  Means of Measuring the Traits 10
  Locating the Subjects 13

III. THE DATA 22
  Test Data 22
  Data on Personality 24
  Data on School Marks 28
  Data on Teaching Records in Training School 30
  Teaching Records in the Field 33
  Measures Expressed in Units of Variability 34

IV. RELATIONSHIP OF THE DATA 41
  The Measurement of Simple Relationship 41
  The Significance of the Several Factors 42
    Age 42
    Size of Preparatory School 43
    "Black Beauty" Records 44
    Extra-Curricular Activities 45
    Mental Tests 49
    Personality 51
    First Semester's Marks 53
    High School Records 54

V. FORECASTING ACHIEVEMENT 57
  Closeness of Relation:
    Simple Correlation 58
    Partial Correlation 60
    Multiple Correlation 63
  Goodness of Prediction:
    Regression Equation for Two Variables 73
    Regression Equation for Three Variables 75
    Regression Equation for Four Variables 79
  Comparison of Levels in Per Cents:
    Single Measures and Achievements 85
    Combination of Measures and Achievement 89

PAGE
VI. THE DISPOSITION OF THE EXTREME CASES 92
Destination of Extreme Variants:
As Shown by Levels of Achievement in Single Measures . 93
As Determined by Regression Equations for Three Variables 97
As Determined by the Regression Equations for Four Variables 99
Origin of Extreme Variants:
As Shown by Levels of Achievement Compared with Positions in Measures 105
As Determined by Regression Equations 109
Justice and Injustice in Elimination:
Of the "Final" Low Grades 115
Of Those Who Dropped Out 119

VII. CONCLUSIONS 123

TABLES

PAGE

I. Comparison of Scores on Cycle Omnibus Test Distributed in Per Cents 14

II. Social Composition of the Group Compared with Other Groups 15

III. Comparison of Training School Facilities 21

IV. Coefficients of Correlations of Age and Other Factors. 42

V. Coefficients of Correlation of Size-Preparatory-School and Other Factors 43

VI. Coefficients of Correlation of "Black Beauty" and Other Factors 44

VII. Coefficients of Correlation of Extra-Curricular Activities and Other Factors 45

VIII. Intercorrelation of the Four Tests 46

IX. Correlations Between the Four Tests and Personality Traits 48

X. Coefficients of Correlation of Mental Tests and Other Factors 49

XI. Intercorrelation of the Eight Traits 50

XII. Coefficients of Correlation of Personality and Other Factors 52

XIII. Coefficients of Correlation of First Semester Marks and Other Factors 53

XIV. Coefficients of Correlation of High School Record and Other Factors 54

XV. Correlation of the Several Measures with the Criteria 58

XVI. Comparison of Relationships Between Measures and Achievements by Means of Partial Correlation . . 62

XVII. Correlation of the Several Combinations of Measures with the Criteria 64

XVIII. Comparison of Relationships Between Measures and Achievement by Means of Simple, Partial, and Multiple Correlation 67

PAGE

XIX. REGRESSION VALUES FOR NORMAL-SCHOOL-SUCCESS ON MENTAL TESTS AND PERSONALITY 78

XX. SIGNIFICANT REGRESSION EQUATIONS FOR PREDICTING ACHIEVEMENT 82

XXI. COMPARISON OF POSITIONS IN TESTS, PERSONALITY AND FIRST SEMESTER MARKS WITH LEVELS OF ACHIEVEMENT . . . 87

XXII. COMPARISON OF POSITIONS IN COMBINATIONS (REGRESSION VALUES) WITH LEVELS OF ACHIEVEMENT 89

XXIII. LEVELS OF ACHIEVEMENT OF THE HIGHEST AND LOWEST TENTHS IN TESTS, PERSONALITY, AND FIRST SEMESTER MARKS 96

XXIV. LEVELS OF ACHIEVEMENT OF THE HIGHEST AND LOWEST TENTHS FOR TESTS AND PERSONALITY; AND TESTS, PERSONALITY AND FIRST SEMESTER MARKS 101

XXV. PREDICTION ERROR FOR THE EXTREME CASES—HIGHEST AND LOWEST TENTHS 103

XXVI. PER CENTS OF THE HIGHEST AND LOWEST TENTHS IN ACHIEVEMENT AS DRAWN FROM THE DIFFERENT LEVELS IN TESTS, PERSONALITY, FIRST SEMESTER MARKS . . . 108

XXVII. PER CENTS OF THE HIGHEST AND LOWEST TENTHS IN ACHIEVEMENT DRAWN FROM THE DIFFERENT LEVELS IN TESTS AND PERSONALITY, AND TESTS, PERSONALITY AND FIRST SEMESTER MARKS 112

XXVIII. PREDICTION ERROR FOR THE EXTREME CASES—HIGHEST AND LOWEST TENTHS—IN ACHIEVEMENT AS TO THE SOURCES FROM WHICH THEY CAME 113

XXIX. ELIMINATION OF FAILURES IN NORMAL SCHOOL WORK ON THE BASIS OF THEIR DEVIATION BELOW THE MEAN . . 116

XXX. ELIMINATION OF THOSE WHO WITHDREW ON THE BASIS OF THEIR DEVIATION BELOW THE MEAN 120

PLATES

PAGE

1. Scores on Tests, Personality and First-Semester-Marks Distributed in Units of Variability 39
2. Scores on Four Types of Success Distributed in Units of Variability 40
3. Levels of Achievement in Two-Years-Marks of Highest 10 Per Cent and Lowest 10 Per Cent 94
4. Levels of Achievement in Normal-School-Success of Highest 10 Per Cent and Lowest 10 Per Cent 94
5. Levels of Achievement in Teaching-in-Training-School of Highest 10 Per Cent and Lowest 10 Per Cent 95
6. Levels of Achievement in Teaching-in-Field of Highest 10 Per Cent and Lowest 10 Per Cent 95
7. Levels of Achievement of Highest 10 Per Cent and Lowest 10 Per Cent on the Basis of Tests and Personality Combined 98
8. Levels of Achievement of Highest 10 Per Cent and Lowest 10 Per Cent on the Basis of Tests, Personality and First-Semester-Marks 100
9. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Two-Year-Marks 106
10. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Normal-School-Success 106
11. Levels Which Supply Highest 10 Per Cent and Lowest 10 Per Cent in Teaching-in-Training-School 107
12. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Teaching-in-Field 107
13. Levels in Tests and Personality Combined Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in the Four Types of Achievement 110
14. Levels in Tests, Personality and First-Semester-Marks Combined Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent for the Four Types of Achievement . . . 111

CHAPTER I

INTRODUCTION

The General Problem of Guidance

With the ascendancy of democracy and the increased complexity of social life, educational guidance has supplanted paternalistic "direction," and the guidance problem is becoming of ever-enlarging scope with correspondingly increased importance and application. Not social and economic status of parent, but individual ability and bent; not "the family tree" and tradition, but neural organization and function; not caste, but character,—the great determiners!

To give expression and guidance to vocational ability and aptitude is at present recognized as a growing obligation on the educational worker everywhere. But the problem of leading the boy and girl into those life activities for which they are best fitted by nature and training, and whose requirements and rewards and possibilities bring satisfaction and contentment and yield to the larger life, is one of recognized importance but difficult of solution. A beginning has been made, and other investigations are in progress, on certain segments of this complex problem. Abilities and interests of elementary[1] and high school pupils[2] and of college students[3] have been studied with the view to determining their educational and vocational fitness,—and with promising results for the immediate future. But little or nothing has been accomplished so far toward the determination of traits and abilities in the normal school student necessary to successful achievement in school and in the class room beyond. That is our particular province and present concern.

---

[1] Kelley, T. L. *Educational Guidance.* Teachers College, Columbia University. Contribution to Education, No. 71.

Fretwell, E. K. *A Study in Educational Prognosis.* Teachers College, Columbia University. Contribution to Education, No. 99.

[2] Proctor, W. M. *Psychological Tests and Guidance of High School Students.* Journal Educational Research Monographs, No. 1.

[3] Colvin, S. S. *Methods and Results of Psychological Tests.* Report to the President of Brown University.

Thorndike, E. L. In *Journal of Educational Research,* May, 1920.

GUIDANCE AND TEACHING

With the growing dissatisfaction with the skill and knowledge and efficiency of its personnel; with the awakened suspicions concerning the quality of the instructional material[1] recruited within its ranks; the time is propitious for Teaching to subject its stock and equipment to mathematical inventory. Waste is becoming more and more apparent. The turnover is unduly large; "help" is loosely selected and poorly placed, with a consequent inefficiency of outcome in educational product. Greater waste with larger numbers in the future is inevitable unless a change in recruiting, training, and placement is effected. Economy will become the handmaid of correct selection of the "would-be-trained" and proper placement of the "product" in the field.

Sound selection of students for the teacher-training work is urgent, even necessary, to the present and near-future situation, if indeed the expansion program for the training of teachers is to be realized. If the normal schools, the institutions chiefly concerned with this professional preparation, are to be elevated to adult college standing, the material with which they deal should be of equal grade to that in the sister institutions. These recruits must possess the ability to make the advance in work accomplished possible, and this ability must be demonstrated if they are to shun the condescensions of those electing the arts and science courses.

Furthermore, if state subsidies[2] to teachers-in-training are to be initiated, it is but just to the public paying the bills that the money be expended economically and to the best advantage; that those for whose preparation it is so expended shall possess real teaching ability; that they have not only the inclination but the ability, not only the attitude but the aptitude.

And if teaching is to be made a profession in reality, as well as in name, the individuals who enter it must possess those abilities and those interests which will enable them to live its ideals, acquire its technique, realize its purpose, and which will insure a high degree of permanence in the work. They must possess not only the right kind of ability and attitude but the right amount—these "writ large." This equipment must be sufficient both in quality and in quantity to give it market value.

---

[1] Yoakum, C. S., and Yerkes, R. M. *Army Mental Tests,* pp. 168–70.

[2] Hertzog, W. S. *State Maintenance of Teachers in Training,* pp. 127–33.

BASIC PHILOSOPHY

That people are different by nature, as well as by training, is more than an assumption; it is a commonplace fact. These differences are not limited to the physical and physiological, but are equally noticeable in matters of intellect and character. They are both qualitative and quantitative,—differences in every human trait varying in amount from imperceptible to enormously large. Some people possess the ability and bent to follow certain lines of related activity and others have different abilities and different bents to follow certain other lines of related activity. Some people are "born short" in respect to certain traits and are necessarily limited in their educational and vocational productivity, and must consequently "find themselves," if they are not to wander aimlessly through life, in that type of activity demanding only such powers as they possess.

That different abilities and interests (the differences being sometimes in kind, always in amount) are required for adjustment to the various activities which life offers; that frank recognition and exact determination of differences in ability, interests, attitude, and requirements for that state of satisfyingness necessary to long-sustained effort in a cause—and permanence in a profession—will make for both efficiency and security; that the best interests of society demand the practical optimum in educational and vocational adjustment of individual and job to secure maximum efficiency with a minimum of waste, these facts are recognized. Consequently, guidance is becoming more and more concerned in scientifically determining relationship and adjustment, in discovering points in common between individual traits and abilities and the vocational requirements and rewards, and in fitting the two together; and, on the basis of work done, progress made, we are encouraged in the possibility of predetermining within reasonable limits the qualities which make for success in the work and in the further possibility of identifying them in individuals who purpose to enter the profession.

Means of tapping the reservoir of human resources and individual possibilities must be sought; means of determining the fitness of the individual for the teaching job must be devised; present practice must be raised from the shifting sands of *a priori* opinion to the plane of scientific measurement. Sound guidance must have its basis in fact instead of in opinion, in science instead of in faith; and justification for conclusions reached must rest finally upon experimental evidence, and future practice upon empirical facts. In proportion to their proved service to humanity in the field are facts of value.

ASSUMPTION MADE IN THIS STUDY

Without serious attempt to set forth elaborately the logic or to mobilize the data to an adequate defense of each position, the following assumptions have been made:

1. That the normal school is primarily and essentially a teacher-training institution, the function of which is to turn out teachers to satisfy present-day conditions and requirements; —not to set standards for administrators in the field, but to supply the product to meet the standards already set, or likely to be set.
2. That the normal-school training course extends over a period of two to four (generally two at present) years of nine months a year; and that the program of studies is composed of a certain amount of academic subject-matter work, a certain amount of methods and professional courses, and a minimum of directed teaching in the training school.
3. That the responsibility rests upon the normal school to "equip" its graduates with the purposes and technique demanded by superintendent, supervisor, and principal, and with certain "social requirements" demanded by the community and parents, and to turn out teachers with those "qualities which school authorities and the world at large *regard* as important." For "Vocational success (*in general, and teaching achievement in particular*) depends not only on the traits one really possesses, but also somewhat on the traits one is believed to possess."[1]
4. That a certain minimum of intelligence is required—the ability reflected in reactions to mental tests, in estimates by instructors and associates, and in school achievement—to meet successfully the school demands and the demands outside of the school.

---

[1] Hollingworth, H. L. *Vocational Psychology*, pp. 166–67.

5. That certain specialized abilities are required, such as those reflected in "trade tests," "professional courses," and "success on the job."
6. That a certain excellence of moral character and ideals is required,—as reflected in estimates of associates, teachers, and employers,—and a certain personality acceptable to those with whom and for whom one works.
7. That human character and individual interests are reasonably stable and consistent, and that what a boy or girl finds suited to his or her tastes and abilities will in a large measure prove suited to the man or the woman, and that life in school should be not merely preparation for the larger life outside but the finding of one's self in that richer and fuller life.
8. So that just in the measure that student-teaching activities in the training school are representative of the teaching activities in the field—and the measuring rod for successful achievement common to both—should training school performance become an index of probable teaching performance outside. To the extent that the former is typical of the latter may we, therefore, expect high correlation and successful prediction.

QUESTIONS INVESTIGATED IN THIS STUDY

Some of the questions to which attention in this study will be given are:

1. To what extent are the abilities and character, the aptitude and attitude, which are necessary to success in the schoolroom, present in the normal-school recruit? And to what extent do they develop during the training period?
2. To what extent do preparatory-school records, estimates of teachers, and test scores identify such traits and abilities in the individual and to what extent do they reflect the amount when so identified?
3. To what extent are traits which are to be essential to teaching success responsible for achievement in school and in the class room? What relation obtains between such personality traits in school and outside?
4. Can prediction of school success be made with any reason-

able degree of reliability on the basis of *a*) test ability, *b*) character estimates, *c*) previous school records?

5. Can prediction of in-field success be made with a reasonable degree of reliability on the basis of *a*) test ability; *b*) character estimates; *c*) high-school records; *d*) training-school achievement; *e*) normal-school success?
6. To which is achievement in school and success outside most closely related,—*a*) test-ability, *b*) personality estimates, *c*) previous school records?
7. To what extent are combinations of these closer to later achievement than the single measures?
8. What is the best expression for this prediction in terms of the partial correlation and the regression equation?
9. To what extent do test-scores, personality-estimates, early school records singly or in "combination" select the "strong" and "weak" recruits?
10. Is there a lower limit which we may safely set for the intelligence of an applicant to the teaching profession? For the personality traits?
11. Is there a "composite lower limit" which we may safely set? One that will insure practical prediction within reasonable limits,—with a not too large probable error?
12. To what extent would it be advisable to reject applicants and to eliminate students on the basis of test results and personality estimates, singly or in combination?

CHAPTER II

THE STUDY AS MADE

During the session of 1919-20 an investigation into the possibilities of predicting the achievement of teachers-in-training, on the basis of data available at the beginning of the year or shortly thereafter, was begun at the State Normal School for Women, Farmville, Virginia. The plan was to assemble past records, measure present abilities and worth, and follow each individual's performance as a student in school and as a teacher outside the first year. Preparatory-school records and character estimates by former teachers were collected (in so far as they were to be had and were in usable form); measurements of intellectual ability by means of psychological tests and personality estimates by instructors at the end of the first semester were procured; the level of performance in school (in terms of marks) and teaching in training school (in terms of ratings) were recorded as the data accumulated; and at the end of the first year as teachers (May-June, 1922) ratings on success were obtained. The present study states the problems as they appeared in the several parts of the investigation; the character, amount, and method of treating the data; the relationship found to obtain between the past and present facts and future success; and the conclusions and recommendations emerging therefrom.

To ascertain the possibility of predicting an individual's achievement as a student in school and as a teacher outside, on the basis of certain measured abilities and traits,—that is, to determine in how far it may be possible to select teacher-training recruits on the basis of available data during the pre-training or early training period,—it was necessary to dispose of three problems at the outset: the What, the How, and the Who. *What* traits should be, or may be, measured? *How* may they be measured; what are the available means for measuring them? *Who* should be measured; who may rightly serve as subjects for the experiment?

TRAITS TO BE MEASURED

Theoretically and ideally the approach would be to picture a sampling of "perfect" or nearly perfect teachers, who had been perfect or nearly perfect students-in-training; analyze their common qualities and characteristics; evaluate and catalogue these characteristics as teacher-traits, and weight in some way the worth of each item in this complexity of traits for achieving the particular type of success; then set out to measure the amount of each such trait present in the applicant; compare the amount thus found in each individual with the amount in the composite ideal; and state in per cents the amount which each recruit possesses. To discover such individual, or individuals, would be difficult, if not impossible; and to measure and evaluate each contributing trait would prove a long and laborious, if not an endless, task.

The method followed to discover what traits count, or rather those considered important,[1] in the teacher's personality, consisted of an examination of reports on studies[2] bearing on the subject, current literature, and opinions of administrators in the field. The lists of traits held to be of importance vary in length, from about a dozen to upwards of a hundred, with no very great agreement as to the fundamentally significant ones. Obviously, it would be impracticable, to say the least, to include all, or even all those thought to be important by a relatively large number of school men. The exercise of choice was a prime necessity, and only those which satisfied the requirements of the study could be selected. Three standards were set up to aid in the selection.

In the first place, the trait to be selected must be as concrete as possible, expressing itself in objective behavior which may be readily observed, manifesting itself in word or act (or both) which is identifiable and unmistakable.

In the second place, it should lend itself to the instruments of measurement available at present. Of two traits considered equally important, the one which is the more sensitive to the means of measurement would be chosen, if only one could be included.

In the third place, the list must not be too long. It must be sufficiently brief that each trait may receive the consideration and attention that its importance warrants.

---

[1] Hollingworth, H. L. *Vocational Psychology,* pp. 166, 167.

[2] Bagley, W. C. *School Discipline,* pp. 30–33.

On this threefold basis, the following abilities and traits were selected to constitute what will be referred to from time to time as Teacher-Personality Traits, or Personality-Traits, or simply Personality (that is, when considered in terms of estimates or judgments which will be explained more fully in the following chapter).

1. Intellectual ability (intelligence, or ability) because of the importance attributed to it as conditioning success achieved and because it lends itself to treatment,—directly by means of mental tests, and indirectly through the school marks.
2. Industry (application, or diligence) because of its considered importance, its relation to school success, and its objective manifestations which are easily noted.
3. Scholarship (academic thoroughness and attainment) because of its known value for predicting school achievement in elementary and high school[1] and because of its measurability (indirectly, at least) in terms of grades or marks.
4. Tact, sympathy, sincerity, optimism, and personal appearance,—by virtue of the importance attached to them by school men and by reason of their manifestations in conduct being fairly marked.

Objections to this list might be made on the ground of its brevity, exclusion of traits equally important, overlapping or commonality of qualities, etc. All of these are, in a sense, valid objections. But to have made the list a long one, of sufficient length to include all important traits, or all those so considered, would have defeated the ends sought; the measurements obtained would have lacked the careful exactness we most desired. No apologies need be made for the overlapping discoverable between desirable traits,[2] provided the correlation be not excessively high,[3] for such is the nature of the human equipment. To secure the coöperation necessary to make the estimates worthwhile, only a comparatively few traits could be included, and those of large and fairly general concern.

---

[1] Kelley, T. L. *Educational Guidance,* p. 81. Teachers College, Columbia University. Contributions to Education, No. 71.

[2] Thorndike, E. L. In *Journal of Applied Psychology,* Vol. 4, p. 27.

[3] Knight, F. B. *Qualities Related to Success in Teaching,* pp. 51-67. Teachers College, Columbia University. Contributions to Education No. 120.

MEANS OF MEASURING THE TRAITS

The second problem in this connection, the determination of means or instruments for measuring these traits, is in large measure complementary to the first,—it places conditions on the selection of traits to be measured and is in turn conditioned in large part by those so selected. When the first problem has been solved as above described, the solution of the second problem follows logically therefrom, and it needs only a brief description and explanation at this point. Practice in the past and present suggests three somewhat separate and distinct ways of tapping human traits and gives promise of results for purposes of this study.

1. *The rating scheme,* by means of which an individual's equipment (traits, abilities, qualities, or what you will) is equated in terms of averages for those in his or her own group.[3] The information does not take the form of a statement that a person possesses, for instance, so much of Trait A, a certain amount of Trait B, and so much of Trait C, which would in the nature of things result in confusion,—since no balances for such "absolute amounts" exist at present, and probably will not at any time in the near future,—but an estimate of a person's standing in Trait A, Trait B, Trait C, etc., as compared with the group. Obviously, it would be impossible to estimate a man's weight in pounds, or his height in inches, if we were ignorant of those units of measurement; but we could say with some degree of confidence (assuming a reasonable amount of experience in seeing men) whether or not he was the largest or smallest, tallest or shortest, just about average, a little above or below average, or very much above or below. So in the absence of scales for human traits, it is reasonable to believe (and to assume) that we can "place" individuals at approximately their proper levels for those qualities which reveal themselves through objective manifestations and with sufficient frequency to be freely observed and clearly recognized. It is more than an assumption; it has been shown to work with varying degrees of satisfaction[1] provided *a*) that the judgments or estimates so given be no less than three in number; *b*) that they be independent one of another, yet have about the same idea in mind; *c*) that people giving them

---

[3] Army Rating Scale. *The Personnel System of the U. S. Army,* Vol. 2.

[1] Rugg, H. O. In *Journal of Educational Psychology,* February, 1922.

be reasonably well acquainted with the individuals being judged; and *d*) that the judges have a reasonable amount of experience with, and knowledge about, the "group" of which the individual judged is a member.

This scheme of individual-to-individual comparison was the one employed in procuring estimates from teachers regarding the "amount" of each of the eight traits present in each student; from the supervisory force on the teaching achievement in the training school as indicated by the items in the "Score Card for Teachers" hereinafter to be described; and from superintendents, supervisors, and principals regarding teaching success in the field, during the first year (using same score card). Moreover, to insure these conditions, a copy of detailed instructions was supplied to each judge, making clear the conditions under which the estimate was to be made,—about what, in whom, how, etc. A minimum of one semester's acquaintance was set, and then only those were rated who were well known to the individual judging. A range of 3 to 7 independent judgments was possible in the first set of measures; 3 to 5 in the second; and 2 to 3 in the third. (For copy of Directions and Score Card, see Chapter III, pp. 24, 25 and 32.)

2. *Mental tests,* by means of which an objective measure of certain types of an individual's intellectual ability may be obtained, seemed to offer much for an investigation such as this, serving as a more or less concrete check on teachers' judgments concerning this same trait. Moreover, they have demonstrated a real value in connection with the problem of prognosis in the elementary school,[1] in high school[2] and in college[3] work. Tests suited to the development of elementary and high school pupils were available in large numbers when this study was begun, but there were very few standardized tests for students beyond the high school level. The choice was much too limited. The following three tests were selected.

a) Thurstone's Cycle Omnibus Test (Psychological Test IV) which was being given at the time to a large number of college and normal-school students, offering, therefore, a basis for comparison.[1] [2] The variety of items included in the test, the circular arrangement of the different types of material, and the increasing difficulty of its parts, make for an interplay (and the exercise) of many functions. Its length (168 items) and difficulty met the requirements of the keenest and swiftest minds, and its less difficult parts suited the needs of the less keen and slower thinkers.

---

[1] Fretwell, *op. cit.,* p. 50.

[2] Proctor, *op. cit.,* pp. 30 ff.

[3] Thorndike, *op. cit.,* p. 331. Also Colvin, *op. cit.* (Report as a whole.)

b) Range of Information Test (devised) of the multiple choice type. Two hundred items of common facts acquired in the course of everyday experience,—observation of environing phenomena, social contacts, current literature, everyday affairs,—not from books or formal school training; information that would become more readily and in far larger measure the possession of the mentally alert than of those forced to live the life of a mental recluse.

The assumption that individuals who quickly become oriented to common life situations possess the alertness which will enable them to identify themselves with the varied and complex problems of the teaching profession, and that by virtue of this alertness, other things equal, they will score higher in the test, and that this score should serve as an index of the heightened probability of future success,—is a reasonable and defensible one. Many embarrassing moments to the teacher, with lessened confidence in his ability, would be avoided by the mere possession of commonplace fact, and many dull moments in the class exercise would be made bright if he (or she) could but vitalize the situation by illustrative materials drawn from his (or her) own resources. Such a test, sufficient in length and in range of items, might well be diagnostic of this ability.

c) *Trabue Language Completion Tests, Alpha and Beta* (Kelley revision), because of the high correlation found to obtain between this type of test and general intelligence as measured by the Binet-Simon group of tests. A certain ability of no mean order is required in this type of reaction,—sagacious guessing of required word or words, plus the feeling and judgment of its correctness and accord with the sense of the whole,—flexible and sensitive sets of experience capable of functioning in novel situations.

---

[1] Thurstone, L. L. In *Journal of Educational Research,* Nov., 1921, p. 276.

[2] Kirkpatrick, E. A. In *School and Society,* January 14, 1922.

3. *School marks* made both in preparatory school (in so far as these were available and in usable form) and in normal school, particularly the first semester's record. Marks are an indirect measure of scholarship, and very roughly an indirect index of ability also, and they serve as a check against the teacher's judgment on this trait. The pupil's school record in the elementary grades has been found to be more closely predictive of the first year's work in high school than either special tests or teachers' estimates of character and ability. Kelley[1] thinks that "Whatever capacity it is that a grade, say, in mathematics, stands for, it is measured with a high degree of accuracy when the records of several years and of several teachers are combined." Whatever qualities and abilities make for procuring good records in school may be found to operate in no small measure in success achieved outside,—especially where life in school makes possible those experiences which harmonize with the individual's abilities, interests, and ambitions, and not only enable him to discover his true self in relation to his fellows, but also afford the opportunities essential to the development of this higher and bigger life. This should be true of the normal-school situation no less than of other types of educational institutions. On this point, only the searchlight of investigation can be depended upon to reveal the facts.

LOCATING THE SUBJECTS

The third part of our present problem is to determine to what extent the 156 students-in-training, whose records and scores and measures of achievement constitute the bulk of the data used in this study, are sufficiently typical of normal-school students generally to insure reasonable validity for the results and conclusions growing out of it; also, to determine to what extent the normal school in which the study was made is enough representative of this class of institutions to make the findings widely applicable. To compare the present group of students with other groups of normal-school students and their school with other teacher-training institutions is our immediate task,—the present purpose of this section.

1. *Comparison with Other Teachers-in-Training.* How do they compare with other teachers-in-training (and with teachers-in-

---

[1] Kelley, T. L. *Educational Guidance*, p. 84.

service in some particulars) with respect to age, educational preparation, mental ability, and social and economic status?

The answer to the first two items, age and scholastic preparation, is that they are typical only for those sections in which the length of the elementary school course is seven years, to which is added four years of high school work. This condition is restricted very largely to the Southern and Southwestern states. They are on the average about one year younger and have had one year less pre-normal school preparation than their fellow students in most of the other states.

Table I compares the mental ability of three groups of normal-school students and one group of college freshmen as measured by the Omnibus Test. (It would be desirable to make similar comparison of the reactions to the other three tests, but data are not available.)

TABLE I

Comparison of Scores on Cycle Omnibus Test Distributed in Per Cents

	Score					Average Score
	38–57	58–77	78–97	98–117	118–137	
Farmville Normal School (First Year)	26.9	43.6	23.7	5.8		70.2[1]
Ten Massachusetts Normal Schools (All Students)	8.6	37.1	34.4	15.1	3.6	75.3[2]
Ten Pennsylvania Normal Schools (Freshmen)	25.0	62		12.3		74.8[3]
Thirty Colleges (Freshmen)	13.2	58		29.0		86.6

[1] Freshmen students only. On 11 years pre-normal school preparation.
[2] All students in normal school. On 12 years pre-normal school preparation.
[3] Freshmen students only. On 12 years pre-normal school preparation.

On the whole, this group tested lower than did the groups in Massachusetts or Pennsylvania. There were noticeably more individuals in the lower ranges and fewer in the upper in the former group (our group)—and none exceeding the score of 117, none in the "A" group according to the above classification—than in the case of either of the other two groups. The difference in the averages, however, is not great.

Since they are one year younger on the average and with one year less school experience than the Pennsylvania group, and two years younger with two years less school training than part of the Massachusetts group, the difference of 5 points less might not be unexpected, and is probably due in large part to the difference in maturity, training, and added experience. Furthermore, the presence of male students in the Pennsylvania group might be expected to increase the scores slightly. The distribution is skew toward the lower end with increased concentration just below the mean. The distributions seem enough alike, however, to suggest that it is a sufficiently representative sample of normal-school population for experimental purposes, in so far as intellectual ability as measured by tests is concerned.

Table II indicates to what extent the students in question are typical of teacher-training recruits (and of teachers generally) from

TABLE II

SOCIAL COMPOSITION OF THE GROUP COMPARED WITH OTHER GROUPS

	Farmville	Missouri	Coffman's
Age at Beginning Teaching	20.3 years	20.4	19.4
Teaching Position—Elementary Grades	69%	56%	
Occupation of Father			
Farmer	43%	48%	44.8%
Professional	8%	12%	7.7%
Business and Trade	20%	28%	14.6%
Artisan	3%		15.4%
Public Official	9%		1.7%
Other Occupations	17%	12%	19.3%
Size of Family	5	6	5
Size of Family Annual Income	$1650 (1919)	$1250 (1914)	$815 (1910)
Brothers and Sisters Teachers	58%	61%	
Fathers and Mothers Teachers	35%		
Members of Family, Normal School or College People	78%	64%	
"Choice Motive"	Love for Work	Promising Job	

the standpoint of social and economic status. Comparison is made with students in the Missouri[1] normal-schools and with teachers in the American public schools in general.[2]

The facts of this table indicate a close correspondence among the three groups in many particulars,—the type of home from which they come, the social and economic level, educational background, type of teaching position, etc.

The sort of individual most frequently found as teacher in the public school is described by Coffman[2] as follows: "The typical American female teacher is twenty-four years of age, having entered teaching in the early part of her nineteenth year when she had received but four years training beyond the elementary schools. Her salary at her present age is $485 a year. She is native born of native born parents, both of whom speak the English language. When she entered teaching, both of her parents were living and had an annual income of approximately $800 which they were compelled to use to support themselves and their four or five children. The young woman early found the pressure, both real and anticipated, to earn her own way very heavy. As teaching was regarded as a highly respectable calling and as the transfer from the school room as a student to it as a teacher was but a step, she decided upon teaching."

The typical teacher-in-training reported in the Missouri survey[1] is pictured somewhat in this fashion. She (based on facts about the women students only) is a little more than twenty years of age, having completed eleven years of elementary and high school education. She is native born with the chances only slightly more than 50 out of 100 that her parents are also American born. Both parents were living when she entered upon teacher-training, and the family annual income—frequently from agricultural pursuits—was approximately $1250, which went to the support of her in school and the five brothers and sisters at home. Her home was in the rural district or in a small town, and her educational preparation was obtained in a 3- to 6-teacher school. Other members of the family (brothers, sisters, mother, father) had attended a normal school (usually that one) and had taught or were teaching at the time. She decided upon teaching as the most remunerative opportunity available. She is undecided as to the length of service in teaching.

---

[1] Bulletin 14, Carnegie Foundation for the Advancement of Teaching.

[2] Coffman, L. D. *The Social Composition of the Teaching Population.* Teachers College, Columbia University. Contribution to Education, No. 41, p. 80.

The Farmville Type. The typical individual in the present group is a little more than twenty years of age upon completion of the training course, having completed seven years of elementary work, four years in high school, and two years in training. Her initial salary is approximately $900 for the school year of nine months. She is native born with an American ancestry of several generations. When she entered upon the training course, both parents were living and had an annual income of approximately $1650, which served to support her at school and the other four children at home. Living in the rural district, she received her pre-normal school preparation in the consolidated elementary and high school located in a nearby village,—three to four teachers in each department, elementary and high school. She early decided to become a teacher, between the years of 13 and 15, her choice being consciously or unconsciously influenced by the members of her family—father, mother, brothers, or sisters, who had taught after attending normal school or college. She looked upon teaching as highly respectable and felt deeply interested in and had a real "love for the work."

The parallelism between the typical American teacher and the teacher-in-training, as measured in the present study, may be indicated thus:

American Teacher	*Farmville Student*
1. Twenty-four years of age; began teaching at the age of 19 plus.	1. Began teaching at the age of 20 plus.
2. Native born of native born parents; five children in family.	2. Native born of native born parents; five children in family.
3. Both parents living when she began *teaching*.	3. Both parents living when she began *preparation for teaching*.
4. Families engaged chiefly in agricultural and mechanical pursuits.	4. Families engaged chiefly in agricultural and mercantile pursuits.
5. Family income of about $800.	5. Family income of about $1650.
6. Began teaching in rural school—thence to town school.	6. Began teaching in "consolidated rural" or small town school of 7 to 10 teachers.

American Teacher	*Farmville Student*
7. Salary $485.	7. Salary $900.
8. From	8. From
"A group least favored and cultural because of its economic status . . . . Represents on the whole an unmarried group which does not perpetuate itself."	A group above the average in cultural elements—an appreciable number of the homes "touched by normal school and college influence."

In the light of the facts detailed above, there can be little doubt of the "average" (or typical individual) of this group being fairly representative of the character of the American public elementary school teacher in general, both in service and in training. A fuller examination of the data shows the parallelism to extend to and include those above and below the "average" as well.

2. *Comparison of Institution with Other Normal Schools.* How does the institution compare with other normal schools—to what extent is it representative of that class of professional institutions? In physical equipment and support? In faculty,—size, quality, and teaching load? In course of study—types of curricula offered? In practice teaching—facilities for and amount required?

The data in the report of the United States Commissioner of Education for 1917 show the Normal School at Farmville (Va.) to be somewhat above the average in value of buildings, grounds, library and laboratory facilities, annual income, enrollment, number of graduates, etc. It was the first institution for the training of teachers established in the State, and the physical equipment has been accumulating along with its history over a period of more than thirty-five years.

The faculty of forty members, including the training school staff, is somewhat above the average for the country as a whole (28) with a relatively large female contingent (75 per cent). In preparation, academic and professional, it proves to be an average normal-school faculty with its due proportion of degrees of the various grades. The median experience is approximately twelve years. The atmosphere and attitude of the place reflect a deep current of professional spirit and purpose common to the members of the group.

The course of study provides the training considered necessary to the proper preparation for the several types of teaching activity

—for the primary, the intermediate, and the high school levels, and for rural and urban centers. An organized group of studies and activities, with a certain amount of overlapping and of common elements, is prescribed for each sub-group—one for those preparing to teach in the kindergarten and first grade, another for those preparing to teach in Grades II, III, and IV, another for those preparing to teach in Grades V, VI, and VII, and another still for high-school teaching. Moreover, especial emphasis is given to rural life problems and to the organization and management of the rural elementary school (primary and grammar grade work) for those preparing to return to the rural districts.

The practice teaching is probably the most crucial factor in the student's preparation for her later work in the field. Investigations have shown that this work is more closely related to the teaching outside[1] than any other part of the normal-school course; others have revealed the fact that, in general, practice teaching is inadequate, both in amount and in quality, to the needs of the prospective teacher;[2] and to realize the values possible from this phase of the training course the following standards have been recommended.

"Practice-teaching conditions should approximate real conditions. To take up the question of conditions first, it is generally admitted that the closer these approximate the real situations as they exist in ordinary public schools the better. This means from 40 to 60 children in a room, divided into not more than two sections. A further condition is the possibility of placing a practice teacher in charge of this situation for a somewhat continuous period—for example, every morning for from 4 to 18 weeks. This does not mean that all the practice teaching in a given normal school need be done under these conditions, since a student may profit a great deal from teaching much smaller groups of children for a half hour a day for several weeks. But to get the *best* results, the more continuous practice under typical school conditions should also be provided."[3]

---

[1] Meriam, G. L. *Normal School Education and Efficiency in Teaching,* p. 80.

[2] Wilson, L. M. *Training School Departments in Normal Schools.*

[3] Judd and Parker. U. S. Bulletin No. 12, 1916, p. 48.

The appended statement from the current bulletin of the school indicates the purpose controlling, and Table III shows how closely the conditions set forth above are approximated in the situation at Farmville.

"The Training School consists of kindergarten, seven elementary grades, and four high school grades; and enrolls about 425 pupils. Any resident, or non-resident child boarding in Farmville may enter, upon the approval of the Training School authorities. The Training School is in charge of a director, supervisors, and heads of departments. A supervisor directs all of the work of a grade, outside of the special subjects, in the kindergarten and elementary grades. In the high school, the departmental plan of supervision is followed, and as a rule each supervisor has charge of a single subject. The director, who is also head of the Department of Education, divides his time between supervision and conducting classes in the Normal School, and helps to link up theory and practice. The Training School enables the students to study the problems of teaching at first hand, and to gain some experience in the instruction and management of children. Students are placed in complete charge of a grade only after they have become familiar with the subject-matter and methods of handling the subjects that they are going to teach; and after they have demonstrated their ability to teach and manage children, through working with individuals and small groups of children, and otherwise assisting the class teacher. Every effort is made to get the students into the work in which they will probably be most successful, and arrangements are so made that the students who are teaching may be practically free for this work.

"No student is allowed to graduate, however proficient she may be in the academic branches, until she has demonstrated that she is qualified to teach, *and that by experience gained in a situation as nearly typical as possible*".[1]

---

[1] Catalog of 1921–1922, pp. 100–101.

Table III below will show in a condensed form to what extent the practice at Farmville approximates the standards set up in respect both to equipment and facilities and to observation, participation and directed teaching in the Training School.

TABLE III

Comparison of Training School Facilities

	Judd and Parker	Farmville
Amount of Practice Teaching per Graduate		
Minimum	100 hours	(96 (I & IV)) (140 (II & III))
Maximum	90 half days	Teaching and Observation
Practice Teaching Facilities—Building		
Trained Director	√	√
Trained Supervisor		
Ratio of Training School Pupils to Student Teachers	100 to 13 (to 30)	100 to 29

Thus, when compared with the teacher-training situation in general, both as to the character and quality of the student personnel and as to the opportunities afforded by the institution for equipping them for service, the representativeness of the Farmville Normal School stands out and shows it to be peculiarly appropriate for a study into the possibilities of applying modern guidance methods to the teacher-recruiting problem.

CHAPTER III

THE DATA*

The data for this study consisted of test scores, personality estimates, school records, and judgments on teaching achievement of practically the entire freshman class who entered the training course, September, 1919.

TEST DATA

The test data were secured early in the year for the four tests from 212 individuals, 156 of whom "completed" the two years' course, 38 others completed one year's work, and 18 "dropped out" with less than a year's work to their credit. The method of administering the tests and the instructions for each were practically the same throughout. Each individual was supplied with a copy of the test and requested to answer as many of the items as she could during the thirty minutes (the tests were given during the regular class period) with the understanding that the test score would have no effect on her class or school record. To avoid disturbing the normal "mental set," the language completion tests (Alpha and Beta) were administered by the teacher of English, the other two by the instructor in psychology.

Scoring the Tests. The scoring was done by the writer more or less mechanically, since the rules were quite rigid and inflexible, permitting but little discretion in any case. On the Cycle Omnibus Test, a credit of 1 was given for each correct reaction with a possible high score of 168. The lowest score made was 40, and the highest 112, the average for the group (156 cases) being a fraction above 70.

On the Range of Information Test, for each item correctly answered a credit of 1 was awarded with a possible score of 200. The lowest score here was 88 and the highest 177, the average for the group (156 cases) being slightly above 137.

---

*A set of the original data is on reference in Bryson Library, Teachers College, Columbia University.

The responses to the Language Completion Tests were graded according to the quality, in accordance with the key, and judged as *wrong* with 0 credit, *partly right* with a credit of 1, and *correct* with a credit of 2. The highest score possible with all omissions correctly supplied was 80 each. The lowest score made on Alpha was 42; the highest, 77. The lowest score on Beta was 50; the highest, 79.

Test-Composite. The nature and brevity of each of these tests prevent them from being adequate measures of intelligence when considered singly. Correlations between each test and every other test and between tests and measures of achievement proved too low for the purposes at hand. To make them a more nearly complete measure of mental ability, the four were merged into a composite by a process of weighting.[1] Accordingly, in subsequent discussions, *tests* will mean this *composite* unless otherwise specifically stated.

The method of merging all into one composite was briefly this. (1) The Standard Deviation for each distribution of test scores was computed. (2) The common denominator of the four S.D.'s (15, 19, 6.6, and 5.8, which is roughly 58) was determined. (3) The respective quotient for each S.D. when divided into this common denominator (giving approximately 4, 3, 9, and 10) was used as a multiplier for each individual score. (4) The individual's "composite score" was the summation of the products of original score in each test times the respective multiplier.

For instance, an individual whose original scores in the four tests were 70, 128, 64, and 63, would have as her composite tests score (70 × 4) plus (128 × 3) plus (64 × 9) plus (63 × 10) equals 1870. And each such composite score might be divided by some constant number to reduce the magnitude, if desired, in the interest of economy of space and to avoid the handling of large numbers. Instead of using the numbers 4, 3, 9, and 10, which are themselves only approximations, the reduction was carried one step further and 1, 1, 3, and 3, were the respective multipliers employed, which yielded practically the same results, and at the same time smaller and more easily manageable numbers. The range was from 460 (lowest) to 723 (highest) with the average at 605.5.

---

[1] McCall, W. A. *How to Measure in Education,* pp. 25-32.

DATA ON PERSONALITY

At the end of the first semester, the teachers were asked for an estimate on each of the eight "personality" traits for those individuals whom they had taught for a half year and about whom they felt confident of being able to judge in those matters. To insure a common interpretation and to eliminate as far as possible individual differences in application of the terms, each trait was defined to mean a certain thing, or a group of things, and each teacher was requested to give estimate on the quality as defined. Furthermore, since the standard was to be the mass judgment for the "average" of the group (that is, a first professional or freshman normal school class and not students in general) precaution was taken against the possibility of comparisons being made with other, and different, groups, by supplying each member of the teaching staff with a copy of the *Directions for Rating Students* detailing the steps to be taken together with an illustration of the method. The instructions given were as follows:

Directions for Judging Traits

1. In rating a student use first-hand knowledge only. Rate only those individuals (First Professional students, in this case) whose work you know pretty well, whom you have taught at least one semester.
2. Assign to each student one mark (or grade) on each of the eight traits. Use no fractions.
3. In estimating each of these traits or qualities use the following plan or scheme:
    - *a.* Select the best student you know (either in this class or in former classes) and mark her "1", as your *standard.* (She may not be rated "1" in every trait, however. It is very exceptional to find a person who deserves to be ranked at the top in all traits.)
    - *b.* Select the poorest (worst) you know (either in this or in former classes) and mark her "9". (She will probably deserve to be rated higher in some of the traits.)
    - *c.* Select the "just average" (either in this or in former classes) and mark her "5". (She will probably be rated higher in some traits and lower in others.)
    - *d.* With these as your *high, low,* and *middle* standards assign to each a grade somewhere from 1 to 9 on each trait; thus:

Illustration of what is meant.

	Miss A.	*Miss B.*	*Miss C.*
Intelligence	1	5	9
Tact	4	3	6
Personal Appearance	6	4	7
Optimism	3	7	4
Sincerity	3	1	3
Scholarship	2	4	7
Sympathy	5	2	2
Industry (Diligence)	2	3	5

Meaning of terms as used:

Intelligence (Intellectual Ability)	What a person is capable of doing or accomplishing; not what one really accomplishes.
Tact	That common sense (good horse sense) which enables one to meet his or her social situations without butting into difficulties.
Personal appearance	Neat and tasty of dress (not expensive), ease of manner, and a well-kept physical self.
Optimism	That quality and attitude of mind of looking on the better and brighter side of things; ability and tendency to see the doughnut rather than the *hole* in it.
Sincerity	That honesty and frankness of mind, freedom from pretense, which leads to speech and act in harmony with ideals and purposes in life.
Scholarship	Attainment in the arts and sciences; thoroughness, as distinguished from superficiality, in courses or studies pursued.
Sympathy	The ability and willingness to put one's self in another's place. Altruistic spirit.
Industry (Diligence)	Attention to, and consistency in, work done. Can be depended on from day to day to be at her post and prepared.

Personality-trait estimates were thus secured on each of the 212 individuals who had been given the tests, all members of the staff (with but few exceptions) who had taught a group of the freshman class contributing to the whole. The judgments on each student ranged from 3 to 7 in number, each judgment being considered equal in value as all the judges were teachers of experience with approximately the same length of contact with students and equally familiar or unfamiliar with the scheme of rating them. The arithmetic average of the several judgments on each particular trait was the individual's "score."

Scoring the Estimates. All scoring and tabulating was done by the writer. In addition to the assumption that all judgments were of equal weight, it was further assumed that equal differences were equal at all points in the rating scale. The distance from 1 to 2 at the high end of the scale was considered the same in magnitude as that from 5 to 6 or 8 to 9. The average, range, and standard deviation computed for the distribution for the eight traits were as follows:

	Lowest	Highest	Range	Average	S. D.
Intellectual Ability	7.0	1.0	6.0	3.7	1.5
Industry or Application	7.8	1.0	6.8	3.3	1.4
Scholarship	7.0	1.3	5.7	3.8	1.4
Sympathy	7.0	1.0	6.0	3.5	1.3
Sincerity	6.5	1.0	5.5	3.2	1.2
Tact	8.0	1.0	7.0	3.8	1.4
Optimism	7.0	1.0	6.0	3.6	1.5
Personal Appearance	6.7	1.0	5.7	3.7	1.4

The different sets of numbers above might suggest lack of discrimination on the part of the judges, that to mark a student high in one trait meant marking her high in the other seven traits, or to mark her low or about average in the one indicated marking her low or about average in the rest. The data show that, while there was a tendency to judge the individual as a whole being,[1] it expressed itself not in the extreme form, probably little more than the real relationship and the common ties between them would warrant.[2] The intercorrelations computed offer not a large amount of evidence of an over-spread of general estimate, which separateness of judgments was probably due in part to the fact that they were asked only to express opinion on the several special and particular traits. They were purposely not made conscious of their estimate of the general worth of the person-to-be-judged, rather the contrary.

Personality-Composite. It was, however, desirable to merge the eight traits into a composite which would afford a better measure of the individual for the purposes of this study than any single measure on any one of the traits. Were they all of equal importance as contributing elements to the teaching-personality, the matter of merging would be one of simple addition,—the arithmetic average of the eight scores,—since all the scores are of like character, and the averages and standard deviations are approximately of equal magnitude. But since they are probably not of equal worth to the teacher's equipment, a weighted average was computed consonant with their relative values as determined by the mass judgment of ten school men experienced in problems of educational practice.

---

[1] Knight, *op. cit.*, p. 52.

[2] Thorndike, E. L. In *Journal of Applied Psychology,* Vol. 4, 1920, p. 27.

They were asked to distribute "100 points" among the eight traits, alloting to each its proportional part as per its contribution to the teacher-personality. The result was a weight of 3 each to Intelligence, Industry, and Scholarship; 2 each to Sympathy, Sincerity, and Tact; and 1 each to Optimism and Personal Appearance —a total of seventeen "shares"—(3×3) plus (2×3) plus (1×2). To obtain the composite score for any individual, it was merely a matter of summing the products of each trait-score multiplied by its assigned weight and dividing by 17. For example, Individual X whose scores in these traits were 3.7, 1.7, 3.3, 1.5, 1.3, 3.3, 2.3, and 5.0, respectively, would be given a composite score of 2.7, i.e.,

$$\frac{3\,(3.7 + 1.7 + 3.3) + 2\,(1.5 + 1.3 + 3.3) + 1\,(2.3 + 5.0)}{17} = 2.7.$$

This scheme was applied throughout the group; a composite score was computed for each student, and a distribution of these combined scores was made. The range was from 1.2 (highest) to 6.2 (lowest),—a distance of 5 spaces as against an average of 6.1 spaces when the traits are considered separately,—with the average at 3.6. This new average and these composite scores will in all subsequent discussions be referred to simply as the *Personality Average,* or *Average Personality,* and *Personality Score* or *Score on Personality,* unless otherwise specifically stated.[1]

---

[1] An attempt was made about the same time to secure like estimates from the teaching staff of the high schools from which the students came; but with limited success, due to the short tenure of teachers in the small schools and to the failure to keep such data as a part of the official record. Only twenty-three cases, all from the same preparatory school, were found to be worth while for statistical treatment, and they will be presented in a

DATA ON SCHOOL MARKS

At the end of the two years' training course, the record for each individual was compiled,—marks made on the various subjects in the Normal School and on their work as student-teachers in the training school. Records were complete for 156 cases (school marks and student-teaching), one year's records only for 38 more, and less-than-a-year's records for the remaining 18 cases. That is, of the 212 individuals who were given the tests, 18 dropped out before the end of the first year, 38 others completed only a year's work, and 156 completed the course.

The Weighted Mean. In computing the average of the school marks (or grades) for each individual, three assumptions were necessary:

First, that the grades or marks given by all teachers and in all departments have equal validity. True enough, there are individual differences in awarding school marks as in all other matters involving human behavior. Some teachers are exacting and rigid in their requirements, others lenient and generous in their grading. Were it possible to get at such constant errors and apply the necessary corrections, the result would be of doubtful value to this type of study which deals in averages based on some twenty-five marks awarded by a dozen different instructors. The chances are many to one that, in practically all cases, the extremes (high and low, plus and minus errors) will neutralize one another, and the obtained average will not be far removed from the true average.

Second, that the difference between any two consecutive marks or grades is the same as that between any other two such marks or grades. That is, the difference between an "A" and a "B" (or "1" and "2", as their numerical equivalents) is the same as that between a "D" and an "E" (or "4" and "5", expressed numerically). We do not know what these differences are, either as to their exact nature or their magnitude. But teachers in general agree that it requires as much more work, or an equivalence in increased quality of work, to raise the mark from "B" to "A" as from "D" to "C", and the reverse. And since the mark awarded for work accomplished is in part, at least, the "product" of the

---

subsequent section as suggestive of the possibilities in the use of that type of data.

teacher's opinion, the assumption of equal differences would seem to be logically well-founded—certainly not a dangerous one.

Third, that any one class-hour of work is the equivalent of any other such unit of the course in so far as the individual's school record is concerned; that any subjects in the curriculum, equal in class-hours per week and in length of term, should have equal weight in the final average,—laboratory and shop periods counting two for one (or one half). The fallacy is apparent, but to prove it statistically would require a discomforting amount of labor and a waste of good paper and pencils. It would not be difficult to demonstrate by means of certain mathematical formulae the greater value of some one subject (say English) as compared with some other subject (say music) in relation to the teaching performance; or of that subject as compared with either one or both of two others (music and chemistry) as related to that end. But to increase the number of subjects to ten or a dozen differing in amounts, each one to be considered in its relation to every other one and to the end (teaching), the problem is multiplied many times, with the reward incommensurate with the labor involved. Dealing in large numbers, the average, weighted according to the "units of work," meets all essential requirements for practical purposes.

Averages for every individual were computed for the First-Semester, First-Year, Two-Years, Academic-Subjects, Professional-Subjects, and other segments of the course, in conformity with those three assumptions. Records were collected, grades "A", "B", "C", "D", "E" were then expressed in their numerical equivalents of "1", "2", "3", "4", "5", class-hours per subject tabulated, and one-semester-hour made the unit. The method may be briefly illustrated by the following case (page 30) in computing the First Semester's average for Miss X.[1]

---

[1] An attempt was made during the early part of the first year to collect the student's preparatory school record (along with character estimates previously noted), but the effort fell far short of complete success, due largely to the dissimilarity of data,—letters, numbers, and per cents. The incomparability of such data rendered them impossible of treatment statistically for the entire group. Twenty-three cases, all from the same school, will be presented in a subsequent section as suggesting a value which such records might have if preserved in a uniform way.

ILLUSTRATION. FIRST SEMESTER'S AVERAGE, ONE INDIVIDUAL

Subject	"Grade"	Mark	Semester Hours	Products
English	B	2	3	6
Psychology	A	1	3	3
History	B	2	3	6
Arithmetic	B	2	3	6
Geography	B	2	3	6
Reading	B	2	2	4
Industrial Arts	C	3	1	3
Total			18	34

Average $\frac{34}{18} = 1.89$

DATA ON TEACHING RECORDS IN TRAINING SCHOOL

Two measurements of the teaching experience in the Training School were obtained,—the composite "grade" of Supervisors and Heads of Departments, and the independent judgments of Director, Supervisor, and Heads of Departments merged into a combined rating. The first was filed in the office of the Registrar and became an official part of the student's school record; the second was solely for the purpose of this study and would in no way affect her graduation or her future work as teacher outside. Comparison of the two confirmed earlier suspicions that "official" records are on the whole more complimentary and flattering to the student than are those yielded in the interest of fact, uninfluenced by the teacher's sympathy and generosity. Moreover, the unbiased judgments yielded higher and more consistent correlations with test scores, personality estimates, and teaching success outside, than did the "official grades," thereby giving additional evidence of error in the latter.

The Estimate Blank. To secure such unbiased judgment, the plan and technique were identical with that employed in connection with personality traits, already described in detail. An estimate card, listing the factors to be judged, together with "Directions for Rating," was supplied to all, and they were asked to "rate" each of their student-teachers on the various items in terms of the performance of student-teachers as a group, and not in comparison with that of experienced teachers. The estimate blank (or card for rating student teachers) came into existence from many sources. The items can be traced to many of the educational articles and to several score-cards widely in use at the time. It was desirable to have a workable series of qualities applicable alike to the teaching situation in the training school and to that in the field. Economy was an important consideration in the selection, it being impossible to include a very large number of items inasmuch as the coöperation of busy and overworked teachers and supervisors had to be secured. Therefore only the most representative items could be included. The organization of the "qualities" followed somewhat closely the practice in the training department at the State Teachers College, Cedar Falls, Iowa. A sample of the blank used is shown on page 32. (Directions were identical in character with those given for Personality Traits on pages 24-25).

Teaching-Score. Again the problem of computing the composite score arose, and again it was met by a system of weightings on the basis of mass judgment. During the session of 1920–21, a group of twenty research students in the Normal School Department at Teachers College, Columbia University, men and women with supervisory and administrative experience, were supplied with a copy of the Rating Card, and requested to "distribute 1000 points among the four main headings and redistribute under each subquality." The same thing was also done by three experts. Tabulations and comparisons revealed a high degree of correspondence between the two sets of distributions, Pearson r of approximately $+.7$; and, on the basis of a combination of these, weights were assigned to the various qualities. The four headings in the order listed on the card were weighted 350 each for the first two and 150 each for the last two; and the specific items in each division were considered equal in value in the scoring, which was not far different from the results obtained by averaging the twenty-three judgments or by taking the average of the two averages.

On the basis of these weights, each estimate blank was scored,—generally three per student-teacher, one from the director, one from supervisor, and another from department head. By virtue of closer contact and greater opportunity for observation, the super-

SERVICE CARD FOR ESTIMATING TEACHERS

Name ______________________	Highest 1	Very high 2	Well above average 3	Slightly above average 4	Just average 5	Slightly below average 6	Well below average 7	Very low 8	Lowest 9	
I. Personal Qualities:										
(1) Ability to meet people										
(2) Self-control and poise										
(3) Promptness and dependability										
(4) Ready command of language										
(5) Cheerfulness										
(6) Sense of humor										
(7) Good judgment										
(8) Initiative and originality										
(9) Accuracy and honesty										
(10) Tact and adaptability										
(11) Fairness										
(12) Force										
II. Teaching Qualities:										
(1) Mastery of the subject-matter										
(2) Selection and organization of facts										
(3) Skill in— *a.* Questioning										
b. Assigning work										
(4) Insight into pupil's difficulties										
(5) Readiness in illustration										
(6) Stimulating pupils to respond										
(7) Success in securing results—										
a. Knowledge										
b. Skills										
c. Attitudes (interests & ideals)										
d. Right habits of study										
III. Managing Qualities:										
(1) Attention to physical conditions										
(2) Economy of time in—										
a. Handling class materials										
b. Routinizing activities										
(3) Control thru directing activities										
(4) Meeting "cases" of discipline										
IV. Community-Force Qualities:										
(1) Coöperation with—										
a. School organization										
b. Community life										
(2) Success in dealing with parents										
(3) Ability to organize										
(4) Willingness to serve										

Remarks:

Why is she not (if not) doing the best work of which she is capable?

1st ____

2nd ____

3rd ____

Supervisor (Supt., Prin.)

____

visor's estimate was considered of most worth and was given the weight of the other two in the final estimate. In computing the final average, it thus became a matter of applying the following formula to each individual case:

$$\text{Composite Score} = \frac{\dfrac{\text{Est. by Director} + \text{Est. by Dept. Heads}}{2} + \text{Est. by Supervisor}}{2}$$

This process was carried through for the 156 students who were given the tests the first year and who completed the two years' training course. The composite scores thus obtained and distributed were found to range from 1.5 (highest) to 6.4 (lowest) with an average at 3.64. In subsequent references to achievement in the training school work or the "Teaching-in-Training-School," these estimates will be meant, and not the "official grades," unless so specifically stated.

Teaching Records in the Field (First Year). The measurements of the first year's teaching experience after graduation were secured in the manner and by the same instrument as those for the training school work,—identical rating cards (with the one omission, "student," from the title) and directions for estimating the various factors (with the one modification that each individual was to be judged in comparison with *teachers-in-general,* and not simply *first-year-teachers*). The scoring of estimates and the computing of averages were also the same. The principal's judgment in this case, however, was considered of most worth, since the nature of his position gives him more frequent opportunity to observe the work of the teachers, and his judgment was assigned the weight of both superintendent's and supervisor's estimates in the final average or score.

Objection to ratings by judges with different standards may be offered, and not without force. It is doubtless true that no two principals have the same experience or the same criteria for evaluating qualities or things more objective,—this being equally true of supervisors and superintendents,—and many of them widely different experience and standards. Such a circumstance is unavoidable in the nature of things,—the teaching force being different in size and quality, differences also conditioned on types of community and local problems, and so on to the end of the vocabulary of differences. But the method employed tends to reduce such differences to a practical minimum. Moreover, the present study is "dealing with a condition and not a theory" as such; and whatever errors creep in because of these differences tend to reduce the correlation, making the true correlation higher than that obtained. So whatever relationship is discovered probably denotes a minimum rather than an average or something higher.

After many attempts and much perseverance ratings were procured on 110 of the 156 original cases,—failure to secure more being due to inability to locate a few of the graduates; a few others were spending the year "at home"; a few others having said *"I will"* in "another venture"; and the rest in situations failing for some unascertainable cause to yield reports and ratings. Moreover, in some of the cases reporting, only two sets of estimates were possible, due chiefly to the absence of the supervisor from the system. In such cases the principal's judgment was given a weight of two thirds.

The combined scores thus obtained yielded an average for the whole group of 3.9 with a range from 1.4 (highest) to 6.0 (lowest), the average being 0.3 of a space lower than that for Teaching-in-Training-School; which is to be expected as many of them were in above-average systems for the state as a whole and consequently judged by people whose standards were correspondingly somewhat above the average. Likewise, they were being compared with others of riper years and of longer experience.

MEASURES EXPRESSED IN UNITS OF VARIABILITY

The individual measurements thus obtained differ in form, magnitude, and significance,—three variations, each one peculiar to the particular means employed. Test scores are relatively large as compared with personality and school marks,—and high scores indicate high grade ability; low scores, ability of a lower level. Personality estimates and school marks are, on the other hand, relatively small (with a possible range of 1 to 9 for the former and 1 to 5 — "A" to "E" — for the latter) with a significance of the reverse order,—the smaller the magnitude of the number, the higher or larger the ability or trait indicated.

For purposes of clarity in understanding and consistency in interpretation and to insure against the confusion consequent on different meanings for numbers of equal magnitude — and of identical meanings for numbers of unequal magnitude — it seems highly desirable that a method be followed which will reduce all measures to a common denominator. Accordingly, any number (40, 60, 75, or what not), when representing an individual's ability or trait, will have the same significance whether applying to tests, personality, or marks; it will indicate rather definitely some one thing or some particular amount.

For the more fundamental purposes of this study, namely, to measure and envisage an individual's position in one function and in groups of functions against the level of achievement in school work and in teaching outside, it becomes imperative that means be employed to translate the different forms into a common language and that nomenclature be adopted which will assist the mental transition from the level in the one function, or group of functions, to the "corresponding" level in other functions or forms of achievement. The need is present to express the "score" or "position" or "standing" in one set of measures (score of 450 in tests, for example) as the equivalent "score" or "standing" or "position" in another set of measures (score of 5.7 in personality) without being compelled to remember that the average for the one (tests) is so-and-so and the average for the other (personality) is so-and-so, amounts very different in magnitude; that the variability for each is a particular but different quantity; that large scores indicate above-average ability in the former but below-average ability in the latter; and so on to the end of logic and arithmetic. To express all measures in the same form will make for economy of time and effort for both the writer and the reader and for their peace of mind.

Such a method exists, enabling us to preserve all individual variations, sound theoretically and not too difficult of understanding or too involved for practical application. The two excerpts which follow state the principle in some detail,—the first as to its validity and the second as to its use.

The first from Woodworth[1] as to its validity:

> There is a way of eliminating both of the troublesome quantities—both the absolute value of the average and the absolute measure of vari-

---

[1] Woodworth, R. S. In *The Psychological Review*, Vol. 19, pp. 99-101.

ability. Let the average in each case be counted as 0; i.e., let the individual's standing be expressed as a *deviation* above or below the average; and further let the measure of variability be taken as the unit deviation, and all deviations be expressed as fractions or multiples of this unit. (For the measure of variability, either the average deviation, or the mean square deviation, or the quartile, etc., may be chosen.) What this method does is to assign each individual's *position in the distribution* of the group: he stands, namely, above or below the group average, and so and so much above or below as compared with the average variation of the group.

No assumption is made by this method as to the ratio between the variability and the group average; for the average is taken as 0 and the variability as 1, the one being independent of the other. The only assumptions underlying the method are those involved in every use of averages and variabilities, namely, that the average means the same thing in respect to one distribution as in respect to another; and likewise, that the measure of variability means the same thing in respect to the different distributions. Both of the assumptions are correct if the distributions are of the 'normal' type, or if all distributions are of any one type. Were one distribution normal, another markedly skew, and a third distinctly bimodal, neither the average nor the average deviation would mean quite the same thing in respect to the three, and the method would be illegitimate; but in such a case it is doubtful if the distributions ought properly to be combined at all. Mental tests usually give group distribution not very different from the 'normal,' though tending on the whole to be somewhat skew in such a way that more individuals lie on the good side of the average than on the bad side. The distributions for different tests do not differ much in shape, and no considerable error can be introduced by placing the average always equal to 0 and the average deviation (mean square deviation, etc.) always equal to 1.

The second from McCall[1] as to its application and use:

The mathematical zero was located at 5 S.D. below the mean instead of at the mean for three reasons. First, this procedure eliminated cumbersome plus and minus signs. Had the zero been placed at the mean, it would have been necessary to report a pupil's (student's) score as *minus 3 S.D.*, or *plus 2.5 S.D.*, or the like. It is much more convenient to report a pupil as 20 or 75, or the like. Second, this procedure gives a convenient range of points between 1 and 100 and locates the reference point at the easily remembered 50. Third, this procedure carries the scale down as far and up as far as anyone will need to go. This would not be the case if the scale ranged only from *minus 3 S.D.* to *plus 3 S.D.* Fourth, this procedure gives a mathematical zero which is close to the supposed absolute zeros for reading, writing, spelling, composition, completion, and other typical mental functions.

---

[1] McCall, W. A. *How to Measure in Education,* p. 294

Thus, to aid in the proper interpretation of scores, whether relating to test ability, personality, school record, or teaching achievement, and to serve as a basis for possible future comparisons, the reference point adopted for all measurements in this study is the mean performance of those recruits who continued in the normal school throughout the two years' training course. Such a reference point can be used in connection with any trait or for any set of measurements; and an individual, whatever his or her ability in any of these functions as measured, will be "placed" at *so-much-above,* or at *so-much-below,* that point. Differences in individual measures are preserved in every case and, at the same time, the several sorts of measures are reduced to a common denominator and comport with a consistent interpretation.

To avoid the troublesome negative (minus) quantities, the reference point (the mean of the group) instead of being referred to as 0 will be 5 S.D. Furthermore, to eliminate as far as possible the fractional forms, this will be expressed as 50 tenths sigma (S.D.), and for purposes of economy of expression called 50 "sigma measures" or simply 50 "units." Thus, any individual's being credited with a score of 50 means that he or she has average ability when compared with the whole group; a score of 40 places him or her 10 units (1 S. D.) *below the average* of the group; a score of 30, just 20 units (2 S.D.) below the average. On the other hand, a score of 60 places one 10 units (1 S.D.) *above the average;* a score of 65 means 15 units (1.5 S.D.) above the average; a score of 72 means 22 units (2.2 S.D.) above the average; and so on. Moreover, a certain score in one function (45 units in tests, for example) means relatively the same position (in this case, five units below the average) or an equal amount of ability in that particular function as this same score in another function (45 units in personality, for example) does for the latter. And 50 in the one is comparable to 50 in the other; 62 in the one is comparable to 62 in the other; and likewise for all amounts of any one trait when compared with the same numerical representations of any other trait.

Only one further point needs to be considered in the transmutation of measures and that is the process of actually computing them from the raw scores. The steps involved may be briefly indicated as follows:

1. An average (mean) for all the individual raw scores in each trait or function is computed.
2. The standard deviation (mean square deviation) for each set of measures is then computed.
3. To obtain each individual's new score—
    - *a.* Subtract his or her *raw* score from the mean (if less than the mean).
    - *b.* Divide this remainder by the Standard Deviation (S.D.).
    - *c.* Subtract this quotient from 50 tenths S.D., which will be the number required.

    (If, however, the *raw* score is greater than the mean, the mean should be subtracted from it and the final quotient added to 50 tenths S.D.)

The resulting measures (tenths-sigma measures or unit measures) will be found to fall largely between 20 and 76 with a tendency to concentrate around the mid-point 50. These so-called unit measures are to be considered as any ordinary measures or scores and should be treated as such in all statistical procedure and in the discussion and interpretation of results. Accordingly, in all later reference to students' scores (or positions, or levels), these tenths-sigma or unit measures are meant unless otherwise specifically stated. A score of 50 should thus be understood to mean average ability; 55 to mean .5 S.D. above the average; 60 to indicate ability 1 S.D. above the average; and so on.

The data for this study enumerated in the first part of the present section are graphically presented in Plates 1 and 2 in the next two pages distributed on the basis of units of variability, in tenths-sigma measures.

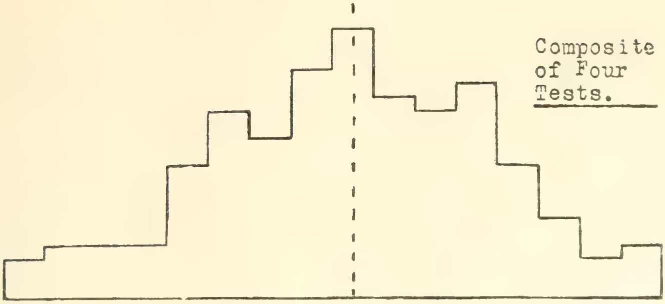


Fig. 1.

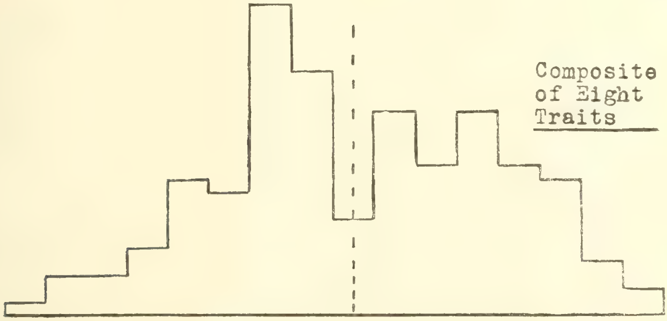


Fig. 2.

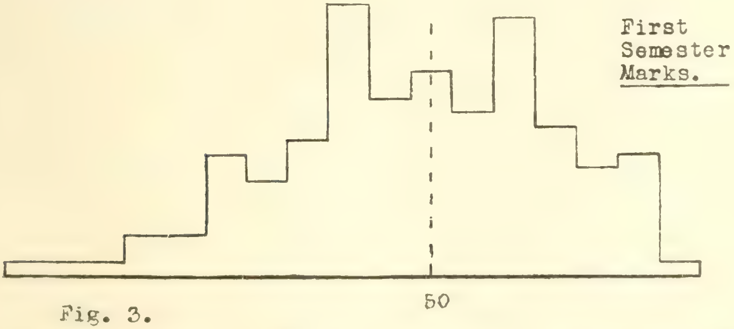


Fig. 3.

PLATE 1. Composite of Four Tests, Composite of Eight Traits, First Semester Marks—Distributed in Units of Variability—Tenths-Sigma-Measures.

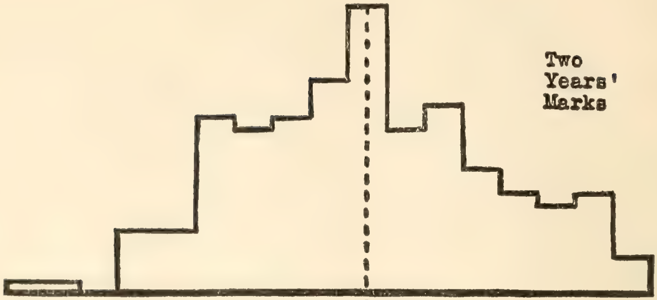


Fig. 4

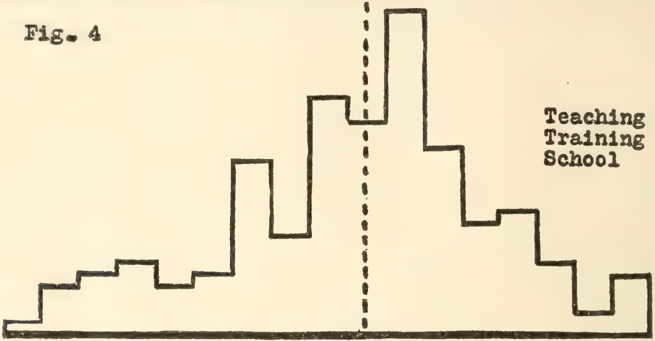


Fig. 5

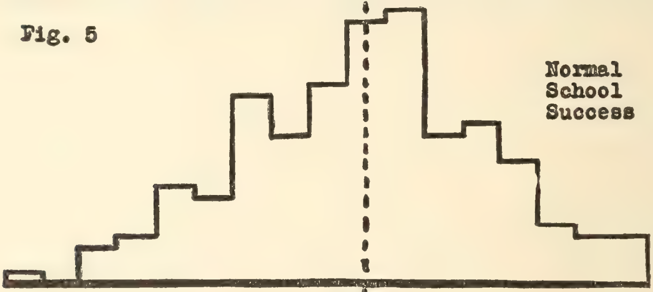


Fig. 6

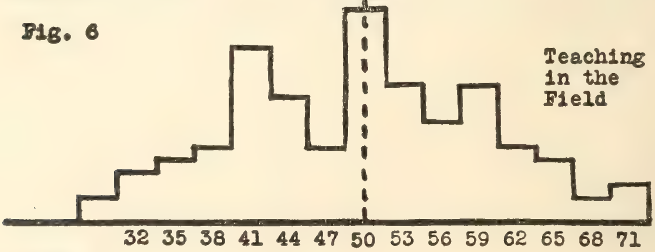


Fig. 7

PLATE 2. Four Types of Success—Two-Years Marks, Teaching-in-Training-School, Normal-School-Success, and Teaching-in-the-Field—Distributed in Units of Variability—Tenths-Sigma-Measures. Average at 50.

CHAPTER IV

RELATIONSHIPS OF THE DATA

The previous chapters have explained the method of collecting the data, the means employed, and the amount compiled. Our present problem is to set forth the relations found to obtain among the several types of data, to show in particular the correspondence between the measures of certain traits and qualities and later achievement with a view to predicting the latter on the basis of the former; that is, to ascertain in how far high position in intellectual ability (or industry, or scholarship, or sympathy, etc.) is associated with high place in school marks, practice teaching, and teaching in the class-room outside; to what extent low position in one or all of these traits is associated with low place in school work and teaching in the field; and to determine the degree of precision possible in stating the one in terms of the other. The "method of best fit" for this twofold purpose is that of correlation[1] and the regression equation.[2]

The Measurement of Simple Relationship

To determine the amount or degree of correspondence between two such series of measures, or the association of the one set of facts with the other set, the Pearson formula $r = \frac{\Sigma(x \cdot y)}{\sqrt{\Sigma(x)^2 \cdot \Sigma(y)^2}}$ has been used throughout. The numerical representation of the "going-together-ness,"—which may vary all the way from + 1.00 (perfect positive correlation) through 0 (no correlation to — 1.00 (perfect negative correlation), generally somewhat removed from either extreme,— will be referred to as the "coefficient of correlation" (and, for economy, "correlation"). The significance of the correlations presented in the following pages will be determined, in general, by the magnitude, since the number of cases and the probable errors are much the same for all: that is, the nearer the coefficient to + 1 (or to — 1), the greater its significance; the nearer to 0, the less its significance.

---

[1] Thorndike, E. L. *Mental and Social Measurements,* p. 173.

[2] Yule, G. U. *An Introduction to the Theory of Statistics,* p. 179.

THE SIGNIFICANCE OF EACH OF SEVERAL FACTORS

The Significance of Age. Coefficients of correlation between achievement in school and age and between success outside and age were computed for 156 and 110 cases, respectively as shown in Table IV.

TABLE IV

COEFFICIENTS OF CORRELATION BETWEEN AGE AND

	**	***
First-Semester's-Marks	.034	± .058
First-Year's-Marks	.023	± .058
Two-Years'-Marks	.051	± .058
Teaching-in-Training School	.117	± .058
*Normal-School-Success	.038	± .058
Teaching-in-Field	.069	± .067

*Measures obtained by taking arithmetic mean of the Two-Years'-Marks and Success-in-Training-School, the two being approximately of equal weight in predicting teaching success outside.

**Pearson r is used throughout to express relationship.

***Probable error is the measure of unreliablity of r.

It will be apparent from these figures that no discoverable relation obtains between the age of these students and the work accomplished by them. In all cases except one, the probable error is equal to or greater than the coefficient of correlation (and to give the latter significance it should be no less than four times the former). No very great importance should be attached to this set of facts since the group was relatively homogeneous in so far as this one factor was concerned,—the youngest being 15 years 5 months on entering upon training, the oldest 21 years 5 months, a difference of six years; the average for the group being approximately 18 years 2 months, with an S.D. of slightly more than one year. We cannot say what the correlation might be, were the individuals more widely distributed in this particular, nor what might obtain for other unlike groups. It is obvious, however, that children of ten to twelve years of age could not make a success in normal school or teaching outside; nor is it likely that adults of fifty to sixty years of age would be able to make the necessary adjustments. It means little more than that "persons" who complete high school between 15 and 21 with the teacher bent have about an equal chance of succeeding in the training work and in teaching outside.

The Significance of the Preparatory School. Correlations were likewise computed for the size of the school in which the recruit was prepared and these various measures of achievement as shown in Table V.

TABLE V

Coefficients of Correlation Between Size-Preparatory-School and

First-Semester's-Marks	.041	±.054
First-Year's-Marks	.059	±.054
Two-Years'-Marks	.086	±.054
Teaching-in-Training-School	−.006	±.055
Normal-School-Success	.038	±.054
Teaching-in-Field	−.049	±.066

It is again apparent that no appreciable relation exists between these. Little or no significance can be attached to the size of the high school (number of teachers per school) from which the students came,—those from the three, four, and five teacher school (high school) doing generally about as high type of work as that done by students trained in the larger city systems. The correlation between the size of the preparatory school and accomplishment is in no case sufficiently high to be significant, and in two cases it shows a slight negative value.

The data are not susceptible of a clear explanation. It may be that the larger systems are, on the whole, superior to the smaller ones but fail to supply from their own student ranks as large a proportion of their superior members to the teacher-training-institutions as do the smaller schools. It is not unreasonable to believe that the relatively better prepared teaching staffs are capable of giving correspondingly better instruction than is possible in the smaller and more poorly equipped centers. Or it may be that the individual attention and acquaintanceship possible (and more widely practiced) in schools where the same instructor comes in contact with the student in two or more subjects, in literary society work, in athletics, and in other extra-curricular activities, compensates for the inferior grade of instruction given. All that we can say at present, however, is that the level of achievement was practically the same for those coming from three- and four-teacher high schools as for those who came from the larger city schools of thirty to eighty teachers and several hundred pupils.

The Significance of "Discipline" Records. The two years' scores from "Black Beauty" (the official register for class-cuts and failure to meet the requirement in classroom work) were compiled and the correlations with accomplishment in school and outside computed. Each deliberate "cut" was given a score of 2; "F" notes (signifying failure in the course) a score of 2 each; and "W" notes (warning that work is unsatisfactory) a score of 1 each. The range in "black marks" was from 0 to 38 (larger scores indicating undesirable records) with an average of approximately 6 (and a median of 3).

TABLE VI

COEFFICIENTS OF CORRELATION BETWEEN "BLACK BEAUTY" AND

First-Semester's-Marks	−.736	±.025
First-Year's-Marks	−.751	±.024
Two-Years'-Marks	−.761	±.023
Teaching-in-Training-School	−.577	±.036
Normal-School-Success	−.647	±.032
Teaching-in-Field	−.527	±.047

All six coefficients of correlation here are fairly high and significant, with relatively small probable error. The "discipline" record so obtained is apparently a very good index of the student's success, and not without reason. Absence from classwork, deliberately incurred, might well be construed by the instructor as evidence of lack of interest with consequent inferiority of work on the part of the student and marks awarded more or less in harmony with this judgment; and the notes sent during the term as warnings of impending disaster, themselves represent judgments-in-the-making and await only the end of the term to be recorded officially as such.

This "influence" is further attested in a negative way by the lower correlations with teaching (both in the training school and in the field). The judgments on teaching outside were wholly independent of this factor, and those on practice teaching largely so; consequently the coefficients of correlation tended to be lower.

The Significance of Extra-Curricular Activities. Measures of the student's activities in other than the regular requirements of the training course,—organization work, committee service, athletics, club life, and the like,—were obtained in terms of "points-per-activity" and a composite score on the basis of weights assigned was compiled for each individual. Identification (membership in and official designation of position held) was the sole criterion of *credit* to be awarded and the *amount* per item was on the basis of mass judgment of their relative worth. The relations to other measures are indicated as follows:

TABLE VII

Coefficients of Correlation Between Extra-Curricular Activities and

Tests	.354	±.047
Personality	.567	±.037
Two-Years'-Marks	.493	±.041
Teaching-in-Training-School	.422	±.043
Normal-School-Success	.501	±.041
Teaching-in-Field	.413	±.053

The correlations between these activities (as measured) and Intelligence, Personality, and Achievement, are appreciable, though not high. They are much too large to be accounted for by chance or mere accident,—the r in every case being more than seven and one-half times its probable error. They signify a real relationship, —probably some common cause which makes for high test scores, high school marks, favorable teachers' estimates, and preferment by one's fellows. Just how close this relationship is we do not know, nor have we the means at present necessary to its discovery. For that a different method and additional data would be required.

The method of awarding *credit* employed here is doubtless open to objections. Obviously, the true worth of an activity is conditioned not on the mere physical presence of the person, but rather on what he or she puts into it in the way of attitude and thought and its outcomes of enlarged possibilities for "further activity."

Careful measurements are required of the investments made and the earned increment therefrom. Furthermore, the judgment of individuals who are participating in the various activities at the time will be necessary to any just evaluation of their relative worth or worthwhileness. It may well be that the amount of energy voluntarily expended has a close relation to its *felt value* on the one hand and to its satisfying outcomes on the other; and these may be widely different from the values assessed by teachers so far, in point of time, from the experience of such activity. The correlations here are at best only a minimum, but sufficiently large to make the subject seem worthy of further and fuller study.

Whatever the validity of the data may be or may not be, whatever the defects in evaluating the activity-items, the relative closeness of relation to the several sets of measures would undergo no radical revision with a change in the method of credits and scores to be awarded. Personality, as judged, would probably yield a higher correlation with voluntary participation in group life than would either sheer intellectual ability or scholarship as measured by school grades. The latter types of mind, as a rule, seem to be more addicted to absorption in abstract ideas and have less time, as well as less inclination, for the multiplication of social contacts on a large scale. For similar reasons, sheer test-ability would seem to be less closely related to such participation and perferment than those qualities and characteristics which make for the winning of "grades" and "marks." Measurements of the real worthwhileness of extra-curricular activities, if such could be had, would probably increase the correlations, though not equally, all along the line.

The Intercorrelation of Tests. The four tests were found to correlate one with another and each with the composite, with and without itself, as follows:

TABLE VIII

INTERCORRELATION OF THE FOUR TESTS

	Omnibus	Information	Alpha	Beta	Composite	Composite Less Self
Omnibus		.483 ±.041	.471 ±.042	.594 ±.035	.750 ±.024	.608 ±.034
Information			.397 ±.045	.472 ±.042	.720 ±.026	.481 ±.041
Completion-Alpha				.782 ±.021	.835 ±.016	.647 ±.032
Completion-Beta					.871 ±.013	.778 ±.021

Tests to be of greatest value for prognosis should yield low correlations one with another, at the same time yielding high correlations with the criteria. Obviously it would be needless, a waste of time and effort, to give two tests or more, which correlate perfectly, both measuring the same function or the same group of functions. The one would serve the purposes of both (or all) with a consequent saving of time and energy. When more than one test is to be used, the ideal,—which is seldom if ever realized,—would be to select tests totally uncorrelated, so as to measure the largest possible combination of traits and functions. This condition could not be attained, nor even approximated, with the tests available for this investigation; but the selection made lacks much of swinging to the opposite extreme. The average intercorrelation between the four tests is .533 with an average deviation of 1.03, and the correlations yielded by the composite with other measures, particularly the several types of achievement we desire to predict, are consistently higher than those obtained from any one test.

Test results to be of value must be consistent: that is, the same test, or tests, must yield in general the same results day after day and with the same type of subject. "Self-correlation is one statistical technique whereby a test's reliability may be determined."[1] The Alpha and Beta Tests consist of similar items approximately equal in difficulty and of known reliability. The observed r between them here is + .782, slightly less than is desired. The reliability for Omnibus Test could not be obtained in this study, but it is reasonable to assume that it would be about the same as that obtained under similar conditions in other investigations with the two series, 1919 and 1920, where it was found that but "few students change their standing very materially,"[2] thus indicating constancy in results. In the absence of two separate Information Tests, the two halves, Part A and Part B, when compared, were found to yield a correlation of + .835, only slightly less than the standard set for individual diagnosis. Taken as a battery, the self-correlations are adequate to insure reasonable validity and reliability.

Mental Tests and Personality Traits. The correlations yielded by each of the four tests and the composite with each of the eight Personality Traits and the composite, were as shown in Table IX.

---

[1] McCall, W. A. *How to Measure in Education,* pp. 396-97.

[2] Kirkpatrick, E. A. In *School and Society,* January 14, 1922, pp. 55-57.

TABLE IX

Correlations Between the Four Tests and Personality Traits

	Omnibus	Information	Alpha	Beta	Tests Composite
Intelligence (Ability)	.469 ± .042	.466 ± .042	.346 ± .048	.578 ± .036	.594 ± .035
Industry (Diligence)	.251 ± .051	.279 ± .056	.209 ± .052	.433 ± .044	.348 ± .048
Scholarship	.457 ± .043	.493 ± .041	.300 ± .049	.537 ± .039	.537 ± .038
Sympathy	.306 ± .049	.299 ± .049	.179 ± .052	.339 ± .048	.347 ± .048
Sincerity	.195 ± .052	.194 ± .052	.100 ± .053	.295 ± .049	.247 ± .051
Tact	.256 ± .051	.162 ± .053	.192 ± .052	.344 ± .048	.307 ± .049
Optimism	.311 ± .049	.298 ± .049	.148 ± .053	.313 ± .049	.322 ± .048
Personal Appearance	.271 ± .050	.360 ± .047	.200 ± .052	.342 ± .048	.371 ± .047
Personality-Composite	.428 ± .044	.412 ± .045	.321 ± .048	.545 ± .038	.517 ± .040

"What a test really measures is known only by its correlations. The closeness of a test's correlation with what constitutes intelligence is a measure of its excellence as an Intelligence Test."[1] Assuming that the teachers really estimated the traits they were requested to estimate, the conclusion is obvious: these tests did measure intelligence and functions closely connected therewith. The correlations with estimated intelligence are appreciably higher than with any other trait except scholarship which is closely connected in thought and in kind; even higher with each of the four tests and with the composite than are the combined traits as a personality composite with the same tests singly and in combination.

Or, reversing the reasoning and assuming that these tests singly or conjointly measure intellectual ability, the estimates purporting to be on intelligence show essentially that the agreement between the two sets of measures is closer than that obtained between the tests and any other set of estimates.

Similar results were obtained when the coefficient of correlation was computed for each of the eight traits (as estimated) with the several segments of success in school and with teaching outside. Scholarship and Ability yielded approximately equivalent *r*'s,—very much larger than any other. Apparently, little discrimination was made between the ability possessed and the scholarship achieved,—*r* between them being + .859; and perhaps but little was possible in the nature of their causal-outcome relationship. One thing is certain: the estimates on those two traits yielded appreciably higher correlations with test scores and school grades, the more objective measures of the traits, than did such estimates on any other traits,—giving added evidence of the validity of the judgments as made.

---

[1] McCall, *op. cit.*, p. 210.

The Significance of Mental Tests. Correlations between the composite of the four tests already described and the several measures of achievement in normal school work and in teaching outside, the first year, were:

TABLE X

COEFFICIENTS OF CORRELATION BETWEEN MENTAL TESTS AND

First-Semester's-Marks	.641	± .032
First-Year's-Marks	.657	± .031
Two-Years'-Marks	.578	± .036
Teaching-in-Training-School	.484	± .041
Normal-School-Success	.540	± .038
Teaching-in-Field	.425	± .052

It is apparent from these correlations, in substantial agreement with a recent study in the county normal schools in Ohio,[1] that intelligence as measured by tests does have a relation, appreciable if not close, to successful accomplishment in teacher-training work and in actual teaching outside. Considering the nature of the training and teaching performances, both highly complex activities, it is clear that many functions other than mere intellectual ability are essential to success, and hence really high coefficients of correlation should not be expected. Particularly is this true when the work is done with a group of highly selected individuals,—graduates of high schools who earlier parted company with most of their fellows less favored with mental alertness. If the maximum correlation expected between tests and college grades be not more than .55 to .65,[2] we should probably not be justified in expecting more than .70 from a slightly less highly selected group of high school graduates who seek entrance into a normal school. These correlations in some cases approximate the maximum hoped for with college students; if they are confirmed by future study with larger numbers, their value for admission and guidance purposes will be difficult of exaggeration.

---

[1] Stevenson, D. H. In *Educational Research Bulletin,* September 1922, pp. 108-111.

[2] Thorndike, E. L. In *Journal of Educational Research,* May 1920, p. 329.

Whether or not this relation be a highly elastic one, stretching out more or less evenly over the entire range, or whether it be such as to sift out the unsuited cases by marking off a lower critical limit below which success is extremely doubtful, if not impossible, may be seen in Chapter VI. Suffice it at present to point out the fact that a correlation exists of sufficient magnitude to be significant.

The correlations thus found indicate a closer relationship between tests and school marks (and higher with academic [.602] than with professional [.486] subjects) than that between tests and teaching, either in the training school or in the field outside. It is reasonable to suppose that school work enlists a relatively larger number of functions involved in reactions to tests than does teaching, and that the latter on the other hand calls into play many qualities, intellectual and temperamental, untouched and unclaimed at present by intelligence tests. In this inference, common sense is reinforced by experimental results.

The Intercorrelation of Personality Traits. The coefficients of correlation between each of the eight traits and every other trait, the composite, and the composite-less-self, were computed with the following results:

TABLE XI

INTERCORRELATION OF THE EIGHT TRAITS

	Intel.	Ind.	Schol.	Symp.	Sinc.	Tact	Optim.	Pers. App.	Pers. Comp.	Personality Composite-Less-Self
Intelligence (Ability)		.604	.859	.574	.476	.532	.599	.604	.868	.798
Industry (Diligence)			.684	.512	.642	.448	.414	.426	.790	.679
Scholarship				.613	.497	.528	.558	.577	.882	.838
Sympathy					.589	.529	.544	.484	.752	.689
Sincerity						.427	.373	.406	.691	.629
Tact							.536	.519	.704	.634
Optimism								.579	.669	.636
Personal Appearance									.678	.635
Personality-Composite										

NOTE: The probable errors range from ± .013 to ± .046, all comparatively small. It is considered unnecessary to include them in this table.

As noted in a former section, the correlation between estimated intellectual ability and observed scholarship is relatively high, which common sense would lead us to expect. Industry and scholarship yield the only other appreciably high coefficient, and that does not seem out of place or contrary to what observation would confirm. The average intercorrelation between the eight traits is + .540, only .007 more than that between the four tests already considered,—certainly not sufficiently large to be alarming.

To what extent the judgment of the individual as a "whole being" affected the judgment concerning a particular trait, or just how far the halo of general effect influenced the particular estimate, the data do not show. Especial care was taken not to bring into consciousness the totality of the person being judged by asking for estimates of general worth as a student, general teaching ability, or the like. On the contrary, nothing was said about the recruit except with respect to the eight specific items or traits. No evidence is present of a "spread of general estimate"[1] unduly large. On the other hand, the correlations in general appear to be little larger than we should ordinarily (reasonably) expect. Moreover, the *differences* between the correlations of specific traits and the composite and that between these traits and the composite-less-self (less each particular trait in turn) are of sufficient magnitude to indicate the absence of any large overflow of "general estimate."

Or, if examination be made from another point of view, it is not difficult to conceive of a correlation between one of these traits and the other seven combined, existing to a point not far below + .69, which is the average obtained from eight such combinations, the range being from + .63 to + .84. Considered individually, the correlations thus found do not appear unreasonably high in view of the obvious relation existing between one or more of the other seven and the particular trait with which they are correlated.

The Significance of "Personality." The following correlations have been found to obtain between the estimated personality (composite of the eight traits explained in a former section) and the records of the 156 students in school and of 110 of them as teachers the following year:

---

[1] Knight, F. B., *op. cit.*, pp. 50-58.

TABLE XII

COEFFICIENTS OF CORRELATION BETWEEN PERSONALITY AND

First-Semester's-Marks	.720	±.026
First-Year's-Marks	.749	±.024
Two-Years'-Marks	.762	±.023
Teaching-in-Training-School	.608	±.034
Normal-School-Success	.720	±.026
Teaching-in-Field	.615	±.040

These correlations are suggestive of a reasonably close relationship between personal traits and educational rewards. Just how close it really is, apart from a certain overlapping influence due to the presence of a judgment common to both sets of measures (with one exception, the estimates on responsible teaching), would be difficult to determine. The mere act of estimating certain of these traits (Ability, Industry, and Scholarship, in particular) in a student doubtless gives the instructor a *set* which will probably influence in some measure the "grade" to be awarded; and each mark given reacts in turn on the teacher's estimate of the individual student.

Fortunately, we have a partial check on this common influence. Although the first semester's marks were awarded by the same teachers who estimated the "personality" traits and although this fact might possibly account for much of the relationship represented by the *r* .720, yet marks for the entire two years were given by many others in addition, many who had no part in judging the traits, and the correlation is even higher, *r* .762, which must be accounted for in other ways. The *r* between Personality and Practice Teaching to the amount of .608 is to be accounted for still less by this common factor, as the "common judges" were relatively less numerous in the two sets of measures; and the *r* between Personality and Regular Teaching, which is slightly higher, .615, cannot be the result of such general factor in the complete absence of the common judge participating.

Nor is it to be explained on the basis of the judgment concerning a common trait. The correlations found to obtain between the individual traits and these six blocks of achievement hold in no one case consistently throughout. The *r's* between Scholarship and the first three (above) are approximately (.747, .760, .755, respectively) the same as those for the eight in composite with those three. But the *r's* between Scholarship and Teaching (both types) are appreciably lower. Estimated ability gives the next best series of *r's* with the several measures of success. No other trait alone even approximates the closeness of relationship discovered for the combination.

Obviously the probable explanation is that those qualities which provoke a high regard by teachers make for success in school, as success is now measured, due allowance being made for a moderate amount of overlapping, and for favorable judgment on teaching achievement outside. And those qualities which have been acclaimed as important teacher traits by school administrators and educational thinkers are manifestations of functions no less important to the equipment of the successful student. In allied or closely related fields the same functions should be expected to operate with marked, though not equal, results, and the data here seem to confirm common sense expectations.

The Significance of First Semester's Marks. Coefficients of correlation between the marks achieved the first half-year and later achievement in the normal school and in the school-room outside were found to be:

TABLE XIII

COEFFICIENTS OF CORRELATION BETWEEN FIRST SEMESTER'S MARKS AND

Second-Semester's-Marks	.759	±.023
Second-Year's-Marks	.695	±.028
Two-Years'-Marks (less self)	.776	±.022
Teaching-in-Training-School	.632	±.033
Normal-School-Success (less self)	.724	±.026
Teaching-in-Field	.600	±.041

The correlations between the First Semester's Marks and other segments of success in school are fairly marked. (The relation of the first to the second semester's work is approximately the same as that obtained by Stevenson,[1] *r* .75, in the county normal schools in Ohio.) The more nearly alike the activities involved, in general, the larger the *r*,—the relation to other groups of "marks" or

---

[1] Stevenson, D. H. In *Educational Research Bulletin*, November 1922, pp. 190-91.

"grades" on subjects being somewhat closer than estimates on teaching-in-training-school. Obviously, the functions enlisted in the service of the latter differ in many particulars, are more widespread and more numerous, from those in the everyday student requirements, and the difference in the magnitude of the *r's* is by no means unexpected.

The *r* with teaching outside lends evidence to support this position. The situation in the field is doubtless more complex, requiring the interplay of relatively more traits and functions, and probably in somewhat larger measure than the practice teaching under close supervision and direction obtaining in the training school. Successful ministrations to and management of the classroom outside demand the display of greater originality, independence, and resourcefulness, than is necessary, or even desirable, during the "apprentice" period; and this relation to their records as students cannot, in the nature of things, be as close. A correlation much higher than *r* .60 with large numbers from many institutions could not well be expected and probably would not be justified, considering the highly selected nature of the group. If this be confirmed by further and more comprehensive investigations, its significance to the guidance problem will be far-reaching.

The Significance of High School Records. The following correlations between (1) the four years' marks received in the preparatory school and the several segments of achievement in the normal school and teaching outside, and (2) the personality estimates by the high school teachers and these types of success in school and out, were computed for the *twenty-three recruits* who came up through the high school department of the normal school and for whom complete records were both available and comparable.

TABLE XIV

Coefficients of Correlation Between (1) H. S. Marks, (2) Personality and

First-Semester's-Marks	.820	±.046	.689	±.074
First-Year's-Marks	.859	±.037	.679	±.076
Two-Years'-Marks	.834	±.043	.730	±.065
Teaching-in-Training-School	.635	±.084	.537	±.100
Normal-School-Success	.808	±.049	.697	±.072
Teaching-in-Field	.774	±.056	.625	±.085

These correlations, with a single exception in the first column and with two in the second, teaching in each case, are comparatively high, and the probable error in every case, except those noted, is relatively small. No explanation to account for the *r's* thus obtained on the basis of chance is possible. That they are as they should be for this particular group seems hardly open to question. Other cause than that of mere accident must be sought to account for the surprising magnitude, if such it be! To what extent this group is typical and representative of recruits in general is, however, of prime concern.

The twenty-three individuals here considered show the characteristics of the whole group studied (the 212 at first or the 156 who completed the course) to a very marked degree. Whatever the set of measurements employed,—Tests, Personality, School Marks, or Teaching,—a few are found at the upper levels, a few at the lower levels, with a larger number concentrated about the mean, and with a high degree of consistency. A few are markedly strong students, and they stand out in contrast to a few others who achieve place at the opposite extreme in practically everything.

The teaching staff and the school itself are, on the other hand, far from being representative in many respects. Unlike the teachers in the typical high school, a staff of three or four teachers who change positions frequently, their tenure is more or less permanent, and in both training and experience they are superior to the "average" high school teacher. Moreover, the official records are preserved complete in the files of the institution, and these are accessible to the teachers at all times. All these facts in themselves make for a wider basis of knowledge and sounder judgment concerning the student.

In how far these conditions, or their equivalent, may be realized in high schools generally and to what degree sound judgments may be approximated, are questions difficult to answer. Attempts to procure such data in the present study proved futile. Suggestions for further, and different, endeavor will be given in a later connection. (See Chapter VII, p. 125.) The importance of data of this type, making due allowance for the more or less ideal conditions—permanent and well-trained teaching staff in a single system—under which they were secured, will be noted in a subsequent section. After all discounts have been made for differences in training, a number of systems, and other inharmonious factors, much of value might still remain if proper methods of taking estimates and sound systems of recording grades were instituted. These are matters by no means impossible of solution, if direction and guidance be given by some centralized agency with authority and initiative to take the necessary steps.

CHAPTER V

FORECASTING ACHIEVEMENT

The present chapter is concerned with the consideration of the predictions possible on the basis of the foregoing data. It is purposed to examine more closely, to analyze the several sets of relationships found, to see if they be such as to prove of value for purposes of predicting the achievement of the student in school and of the teacher outside; and to ascertain which are most valuable and which least valuable for this purpose; furthermore, to determine the possibilities of combining certain of the measures to increase their predictive worth,—increased prediction with reduced error.

The immediate problem is, therefore, a twofold one:

First, to compare the relationships,—both apparent and real,—which have been found to exist between Test-ability, Personality, and First-Semester-Marks and Accomplishment. To answer the questions,—What tests or measures are of value,—are appreciably, and in reality, related to success? Which are of most value,—most closely related to success? What combination, or combinations, may be effected to insure increased value,—to improve this relationship?

Second, to determine the relative "goodness of prediction" of the several measures. That is, to answer the questions,—Which set of measures, or combination of such measures, will give the "success" level of individuals over the entire range with the least error? What is the best form (or forms) for expressing this prediction, and how may they be applied and used?

The Closeness of Relation

In answering the first part of the problem, the closeness of relationship, we must examine the correlations found to obtain between them,—the comparative worthwhileness of the several sets of measures (Tests, Personality, First-Semester-Marks, etc.) being determined on the basis of their correlations with the facts they purport to predict;[1] i.e., success as students in school and as teachers in the class-room outside. That is, the value will be in terms of observed relationship, expressed as coefficients of correlation, and in proportion to the closeness of its relation to the several segments of success which it is used to forecast.

The first test applied here will be a comparison of the correlations of each set of measures with each sort of achievement disregarding the presence or absence of other factors which may possibly influence the observed relation. In Table XV are assembled the more important correlations enumerated in Chapter IV between the different measures and the four criteria. The magnitude of the *r* in each case indicates the relative closeness of the relation since the numbers remain constant (156 and 110) throughout and the probable error small,—equal validity for the several measures being assumed.

TABLE XV

CORRELATION OF THE SEVERAL MEASURES WITH THE CRITERIA COMPARED TO SHOW THE RELATIVE CLOSENESS OF RELATION

Measures	Two Years' Marks (n=156)		Teaching-in Training School (n=156)		Normal School Success (n=156)		Teaching in Field (n=110)		Average
Intelligence Tests.	.578	± .036	.484	± .041	.540	± .038	.425	± .053	.507
Personality Estimates	.762	± .023	.608	± .034	.720	± .026	.615	± .040	.676
First-Semester's-Marks	.776	± .021	.632	± .033	.724	± .031	.600	± .041	.680
Teaching-in-Training-School							.700	± .033	.753
Two-Years'-Marks							.707	± .032	.756
Normal School Success							.734	± .030	.734
"Black Beauty"	-.761	± .023	-.577	± .036	-.647	± .032	-.527	± .047	-.638
Extra-Curricular Activities	.493	± .041	.422	± .044	.501	± .040	.413	± .053	.457

In the upper part of the table are given the facts which are obtainable during the early part of the training course; in the lower part, those which are accumulated during the entire two years' period. The former may be used to predict both in-school and in-field achievement, the latter for teaching in class-room outside only. The measures are here listed in the order of their magnitude of relation,—Tests, Personality, and First-Semester-Marks,—with the four classes of achievement as per the average of the four *r*'s in each case. The same holds for the individual *r*'s with a single exception, with teaching-in-the-field, in which case Personality yields a correlation slightly higher than that found for First-Semester-Marks. There is little doubt that any material displacement would result in the order here indicated, should other measures be taken under similar conditions and treated in a similar fashion. The probability is that the coefficients of correlation would remain virtually as they are since the probable error ranges from ± .02 to approximately ± .05.

---

[1] McCall, W. A., *op. cit.*, p. 210.

Tests show correlations from somewhat above + .40 with Teaching to slightly less than + .60 with Two-Years'-Marks; Personality yields *r*'s approximately two-tenths higher generally, or slightly above + .60 and a little less than + .80, respectively; and First-Semester-Marks practically the same as Personality. The order of importance of these,—with the chances about 50-50 between the second and third,—is clearly apparent. Tests yield the least close relation, the lowest correlation, of the three.

While the order of the apparent relationship is clear, the real relationship may be questioned,—may prove to be somewhat different. Test ability, when compared with the other two in relation to either success in school or teaching outside, without considering the effect of other influencing factors, is comparatively low. Just what effect other operating factors have may not be known directly from the above numbers themselves. When properly treated, any one of three possible results may occur,—differences thus found may disappear, they may be accentuated, or they may remain practically as they are.

We do not know whether test-ability, *per se,* helps or hinders a student in acquiring school marks as a student and favorable estimates as a teacher. We do not know whether personality, in and of itself, be a handicap or an aid in making a good school record and a good teaching record. We cannot say whether the school record made the first half year, apart from other factors, be an asset or a liability in later endeavor in school and in the classroom outside. Nor can we say from these facts alone how close the relation between intelligence and achievement would be if personality or school marks (or both) were not operating at the same time. Nor can we tell offhand how close or how slight the relation would be between each of these two and success, were it not for the influence of one or both of the other two factors. The data show that those who reacted best to the tests were as a rule judged to be highest in personality traits, and they made the best school records the first half year, as well as the greatest success later. But they do not show whether they were able to achieve more *because* of a superior intellectual ability or *because* of the better personality and a good start in school work, one or both. Concomitance is a fact; causal-effect relationship may be questioned.

Furthermore, the correlations show that the better teachers have more intelligence, more often have stronger personality, accumulate higher school marks, win fewer "Black Beauty" scores, take a larger and more prominent part in extra-curricular activities; but they do not show whether those individuals who finally achieve the greatest success in teaching do so because of superior intelligence, personality, school marks, "Black Beauty" record, or voluntary participation in group life.

If we were considering a group of teachers who were equal in all essential particulars except one,—alike in Personality, School Records, etc., differing only in Intelligence; or alike in Intelligence, School Marks, etc., differing only in Personality; or alike in Intelligence, Personality, etc., differing only in School Records,—cause could be argued with respect to the particular factor which varied with the success or achievement. Such groups of teachers are impossible to locate; no one factor can be thus isolated and shown to be the cause of the difference in the levels of achievement. In the absence of such life conditions, a more complicated statistical technique than simple correlation is necessary.

Partial Correlation. Partial correlation is the technique by means of which any single factor may be isolated, freed from its association with any number of other factors. That is the method employed here to determine the real relationship of the several factors to the different forms of achievement, and to determine the "independent" contribution of each one to the predicted success,—that is, to find out just how closely ability-to-make-a-high-score-in-tests is related to ability-to-achieve-high-position-in-school-work or teaching-in-class-room-outside, when the factors of Personality and First-Semester-Marks (one or both) are held constant; likewise for each of these when the other two (one or both) are made non-active.

The formula[1] is:

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{(1-r^2_{13})(1-r^2_{23})}}$$

in which $r_{12\cdot3}$ signifies the correlation between factors 1 and 2 for a constant value of factor 3, when the effect of factor 3 is neutralized. This formula may be expanded and applied so that any number of factors may be treated, so that the relation between any two factors, freed from their connection with any number of other factors, may be determined. We have found it unnecessary to carry the treatment beyond three and four variables.

Table XVI below shows the more significant relations found to obtain between each set of measures and each of the four classes of achievement when one or both of the other related factors have been rendered constant or inoperative. Three partial correlations between each factor (Tests, Personality, and First-Semester-Marks) and each type of achievement (Two-Years'-Marks, Teaching-in-Training-School, Normal-School-Success, and Teaching-in-Field) have been computed; one each in which each of the others is out, a third in which both of the other two factors have been neutralized. An average for the first two is drawn, and that in each case indicates the relation which exists when one half of the influence of each of the other two factors in withdrawn. For example, a correlation of r .332 between Test-Ability and Two-Years'-Marks, with Personality constant, r .167 with First-Semester-Marks constant, r .249 with one half of each withdrawn, and r .144 when both are rendered non-operative. That is, with Personality equal, there is a residual correlation between Intelligence, as measured by Tests, and success in accumulating grades in the two years' course of +.332; with First-Semester-Marks equal, still an r of +.167; and with both of these equal, it is +.144. The amounts of reduction are +.246, +.411, and +.434.

The data in the lower part of Table XV, regarding Extra-Curricular Activities and "Black Beauty" records, were accorded similar treatment, with results which would not warrant further consideration for present purposes, nor could their inclusion in subsequent tables be justified on the basis of real worth. Extra-Curricular Activities when correlated with actual teaching, with Personality held constant, yielded only +.091; with Two-Years'-Marks equalized, +.105; and with Normal-School-Success out, +.079. "Black Beauty" record and Teaching, with First-Semester-Marks held constant, yielded a correlation of +.157; with Two-Years'-Marks out, –.002; and with Normal-School-Success equalized, +.008. Had the process in every case been carried to the second order instead of only to the first, rendering two factors constant instead of one (both Personality and School Marks, for instance,) the residual correlations would have been further reduced. But this was not considered necessary in view of the smallness of these independence values. The contribution of Extra-Curricular Activities to Teaching, apart from that carried by Personality, is negligible; and that added by the "Black Beauty" record, apart from what the Two-Years'-Grades give, is practically zero. The former is doubtless tied up very closely with the personal qualities of the individual; the latter serves as a sort of barometer to register the instructor's judgment of the student's scholastic stress periods.

---

[1] Yule, G. U. *An Introduction to the Theory of Statistics*, pp. 238.

TABLE XVI

Comparison of Relationships Between Measures and Achievement by Means of Partial Correlation

	Two-Years' Marks- (*M*)	Teaching-in-Training-School (*T*)	Normal-School Success (*N*)	Teaching-in-Field (*F*)	Aver.
Tests (1)	$r_{M1\cdot 2} = .332$	$r_{T1\cdot 2} = .250$	$r_{N1\cdot 2} = .282$	$r_{F1\cdot 2} = .158$	.256
	$r_{M1\cdot 3} = .167$	$r_{T1\cdot 3} = .133$	$r_{N1\cdot 3} = .143$	$r_{F1\cdot 3} = .066$	.127
	.249	.192	.213	.112	.192
	$r_{M1\cdot 23} = .144$	$r_{T1\cdot 23} = .109$	$r_{N1\cdot 23} = .113$	$r_{F1\cdot 23} = .034$	.100
Personality (2)	$r_{M2\cdot 1} = .663$	$r_{T2\cdot 1} = .478$	$r_{N2\cdot 1} = .612$	$r_{F2\cdot 1} = .510$	.566
	$r_{M2\cdot 3} = .465$	$r_{T2\cdot 3} = .284$	$r_{N2\cdot 3} = .615$	$r_{F2\cdot 3} = .330$	.374
	.564	.381	.514	.420	.470
	$r_{M2\cdot 13} = .456$	$r_{T2\cdot 13} = .275$	$r_{N2\cdot 13} = .410$	$r_{F2\cdot 13} = .325$	.367
First-Semester-Marks (3)	$r_{M3\cdot 1} = .647$	$r_{T3\cdot 1} = .479$	$r_{N3\cdot 1} = .585$	$r_{F3\cdot 1} = .472$	.546
	$r_{M3\cdot 2} = .506$	$r_{T3\cdot 2} = .352$	$r_{N3\cdot 2} = .427$	$r_{F3\cdot 2} = .288$	.393
	.577	.416	.506	.380	.470
	$r_{M3\cdot 12} = .425$	$r_{T3\cdot 12} = .276$	$r_{N3\cdot 12} = .348$	$r_{F3\cdot 12} = .247$	.324

The facts above add confirmation to the inferences drawn from Table XV and show furthermore that a residual relation exists when one or both of the other related factors have been equalized or rendered non-operative. In no case is the value of one set of measures "swallowed up" by the presence of another set, nor has the relation entirely disappeared by removing the influence of the other factor, or factors. The order of closeness of relationship remains virtually the same, but the amount has, in all cases, been reduced,—moderately in some cases, much in others.

Personality yields a correlation with all four classes of success, with the exception of Teaching-in-Training-School, slightly in excess of that for school grades received the first semester, with an average of only .043 higher, not sufficient to be really significant. Test-ability proves to be less highly correlated in every instance, but in all cases a positive correlation is found to exist. Undoubtedly, all three factors have predictive value, the exact amount to be indicated in the second part of the present chapter.

We have found for no one of these factors a sufficiently high correlation with any one of the four types of success to serve as an adequate basis for individual prediction.[1] At the same time we have shown a real relation to exist between each factor and each sort of achievement apart from the influence of other factors. No one factor alone suffices for our purposes. A method is needed which will enable us to add the correlation between one factor (or set of measures) and success to that between another factor and the same success, not already accounted for in the first; and this to a third not already accounted for in the first two. That is, we need to add the "independent" correlations to that which is common, so as to boost the simple, or crude, correlations.

Multiple Correlation. We shall probably never find any single factor, or set of measures, which will reveal 100 per cent (or perfect) relationship to achievement in school or to teaching outside. These activities are complex and many-sided, involving numerous traits and functions. Ability to perform in tests is doubtless one; personality is unquestionably another; a certain store of accumulated facts another; and probably there are many others untouched in this study. Types of performance such as these can be explained only on the basis of complex causation, and for its expression,—

---

[1] Thorndike, E. L. In *School and Society*, February 15, 1919.

that is, correlation between several contributing factors or variables, —multiple correlation is required. The relation when so expressed is the maximum correlation obtainable between one factor and all others in the combination properly associated each with every other one.

The formula[1] is:

$$\sigma^2_1(1-R^2_{1\cdot23}) = \sigma^2_{1\cdot23} = R^2_{1\cdot23} = 1 - \frac{\sigma^2_{1\cdot23}}{\sigma^2_1}$$

$$R^2_{1\cdot23} = \sqrt{1-\frac{\sigma^2_{1\cdot23}}{\sigma^2_1}}$$

in which $R_{1\cdot23}$ indicates the correlation between factor 1 and the combination of factors 2 and 3. This formula may be expanded to include any number of factors. We have not found it necessary to consider a larger number than four at a time,—the one correlation with a combination of three others ($R_{1\cdot234}$).

Table XVII below shows the correlations thus computed between the different criteria of success and these factors in combination, two-at-a-time and three-at-a-time.

TABLE XVII

Correlation of the Several Combinations of Measures with the Criteria Compared to Show their Closeness of Relation

Multiple R

Combinations	Two-Years'-Marks ($n=156$)	Teaching in Training School ($n=156$)	Normal-School-Success ($n=156$)	Teaching in Field ($n=110$)	Aver.
Tests and Personality	.792 ± .020	.640 ± .032	.746 ± .024	.628 ± .039	.702
Tests and First-Semester's-Marks	.783 ± .021	640 ± .032	.732 ± .025	.603 ± .041	.690
Personality and First-Semester's-Marks	.828 ± .018	.676 ± .030	.778 ± .022	.655 ± .036	.734
Two-Years'-Marks and Teaching in-Training-School				.740 ± .029	
Tests, Personality, and First-Semester-Marks	.833 ± .017	.680 ± .030	.781 ± .021	.658 ± .036	.738
Tests, Personality, and Normal School-Success				.748 ± .028	

---

[1] Yule, G. U., *op. cit.*, pp. 247 ff.

If one compares the facts of Table XVII with those in Table XV, it is evident that combinations of measures yield higher correlations generally than do single measures. This is true for groups of factors themselves positively related to the factor or variable with which the combination is correlated. Over against *r's* of .587 and .762 for Tests and Personality, respectively, with Two-Years'-Marks, the combination yields an *r* of .792; as against *r* .484 and *r* .608 for these with Teaching-in-Training-School, the two together produce a correlation of *r* .640, an appreciable increase; and so on for other possible comparisons.

To indicate the contribution of Tests, Personality, and First-Semester-Marks each to the "combined" correlation, the average was computed for each of the four pairs of coefficients to which it is one of the contributing factors. The averages yielded with the four divisions of success were as follows:

Tests	.788,	.640,	.739,	and	.615,	respectively
Personality	.810,	.658,	.762,	and	.641,	respectively
First-Semester-Marks	.806,	.658,	.755,	and	.629,	respectively

Again it is evident that since tests are less closely related to achievement, they make relatively smaller contribution to the combination than either of the other two,—with no appreciable difference in the added increments by personality and early school marks. Tests, Personality, and First-Semester-Marks combined into a conjoint measure yield correlations not appreciably higher than two of them so combined,—only slightly larger in most instances. If, however, Normal-School-Success be substituted for First-Half-Year's-Grades, the *r* between them and actual teaching is markedly increased,—from .655 to .748. Tests and Personality together are about as closely related to the complete school record as these together with the school record are related to the measured teaching performance in the field.

Simple, Partial, and Multiple Correlation. The more important facts of Tables XV, XVI, and XVII, are summarized in Table XVIII (page 67). The coefficients for partial and multiple correlation for three factors, or variable ($r_{12 \cdot 3}$ and $R_{1 \cdot 23}$) therein represented are averages based on actual coefficients to the number indicated in the parentheses following. All others in the body of the table are actual coefficients as computed. The average for each type of coefficient expressing the relation between a set of measures, one factor (Test-ability, for example), and the several types of achievement, are shown in the column at the extreme right. For instance, for the four crude *r*'s between Test-ability and Success in Two-Years'-Marks, Teaching-in-Training-School, Normal-School-Success, and Teaching outside, the average is .507; for the four partial $r_{12 \cdot 3}$'s, each being itself the average of two (one of which is the correlation between Tests and Achievement with Personality rendered constant, the other with First-Semester-Marks equalized) representing the closeness of relation between Test-ability and each type of success when one half of the influence of the other two factors is withdrawn, the average is .192; for the four partial $r_{12 \cdot 34}$'s, from which in each case the entire influence of the other two factors has been withdrawn, the average is .100; for the four multiple $R_{12 \cdot 3}$'s, each being in turn the average of two coefficients (in one of which Test-ability is "combined" with Personality, and in the other with First-Semester-Marks), the average is .696; and for the four multiple $R_{12 \cdot 34}$'s, in each of which all three factors are "combined" in the correlation with success, the average is .738; and similarly for the rest of the table.

The series of averages at the bottom of the table apply in like manner to the types of success and indicate the relative predicableness of the four classes of achievement.

TABLE XVIII

COMPARISON OF RELATIONSHIPS BETWEEN MEASURES AND ACHIEVEMENT BY MEANS OF SIMPLE, PARTIAL, AND MULTIPLE CORRELATION

	Type of r	Two-Years'-Marks	Teaching-in-Training-School	Normal School-Success	Teaching-in-Field	Av.
Tests (1)	r_{Z1}	.578	.484	.540	.425	.507
	$\frac{r_{Z1\cdot2}+r_{Z1\cdot3}}{2}$	.249 (2)	.192 (2)	.213 (2)	.112 (2)	.192
	$r_{Z1\cdot23}$	.144	.108	.113	.034	.100
	$\frac{R_{Z\cdot12}+R_{Z\cdot13}}{2}$	.788 (2)	.640 (2)	.739 (2)	.616 (2)	.696
	$R_{Z\cdot123}$	.833	.680	.781	.658	.738
Personality (2)	r_{Z2}	.762	.608	.720	.615	.676
	$\frac{r_{Z2\cdot1}+r_{Z2\cdot3}}{2}$	.564 (2)	.381 (2)	.514 (2)	.420 (2)	.470
	$r_{Z2\cdot13}$	.456	.275	.410	.325	.367
	$\frac{R_{Z\cdot12}+R_{Z\cdot23}}{2}$	.810 (2)	.658 (2)	.762 (2)	.642 (2)	.718
	$R_{Z\cdot123}$	,833	.680	.781	.658	.738
First-Semester Marks (3)	r_{Z3}	.776	.632	.724	.600	.680
	$\frac{r_{Z3\cdot1}+r_{Z3\cdot2}}{2}$	.577 (2)	.416 (2)	.506 (2)	.380 (2)	.470
	$r_{Z3\,12}$	.425	.276	.348	.247	.324
	$\frac{R_{Z\cdot13}+R_{Z\cdot23}}{2}$	.806 (2)	.658 (2)	.755 (2)	.629 (2)	.712
	$R_{Z\cdot123}$	.833	.680	.781	.658	.738
$1/3\ (r_{Z1}+r_{Z2}+r_{Z3})$		.705 (3)	.575 (3)	.661 (3)	.547 (3)	.662
$1/6\ (r_{Z1\cdot2}+r_{Z1\cdot3}+r_{Z2\cdot1}+r_{Z2\cdot3}+r_{Z3\cdot1}+r_{Z3\cdot2})$		.463 (6)	.330 (6)	.411 (6)	.304 (6)	.377
$1/3\ (r_{Z1\cdot23}+r_{Z2\cdot13}+r_{Z3\cdot12})$		.342 (3)	.220 (3)	.290 (3)	.202 (3)	.263
$2/6\ (R_{Z\cdot12}+R_{Z\,13}+R_{Z\cdot23})$		.801 (6)	.652 (6)	.752 (6)	.629 (6)	.709
$3/3\ (R_{Z\cdot123})$		.833	.680	.781	.658	.738

NOTE: The subscript z refers to the particular kind of achievement with which the measures (Tests, Personality, and First-Semester-Marks) are correlated; the subscripts 1, 2, 3, refer to the *measures* themselves as indicated in the parentheses under each. For example when r_{Z1} denotes the correlation between *Two-Years'-Marks and Tests* it is .578; when it denotes the correlation between *Teaching-in-Training-School and Tests* it is .484; and so on. The z here takes the place of M in the first column of coefficients (See Table XVI); of T in the second column; of N in the third; and of F in the fourth. This holds true for all three types of correlation,—simple, partial, and multiple.

A comparison of these coefficients of correlation (simple, partial, and multiple) for each factor (Tests, Personality, and First-Semester-Marks) with the four types of achievement leaves little doubt as to the relative "closeness" of these measures to success in school and in the field. Test-ability yields consistently correlations smaller in magnitude than either personality or first-semester marks. And early school marks reveal slightly closer kinship to later scholastic achievement than does personality. These two yield about the same r's with practice teaching, but personality is a little closer to responsible teaching the year following.

The explanation seems clear. It is not unreasonable to believe that the same factors which operate to win high marks during the early months of normal school life continue effectively to so operate throughout the two years. (The *r's* between First- and Second-Semester's-Marks and First- and Second-Year's-Marks, .759 and .708, respectively, would seem to support this view.) The activities involved in study being of the same sort in general from beginning to end, the teaching staff remaining relatively constant, the functions which make for success at first, whatever they be,—many or few, large or small, strong or weak,—doubtless persist with consistent results. However, there are always a few individuals in every student-group who begin well but end poorly, and a few who begin poorly but end well, whose "records" place them in the category of exceptions. But there are usually discoverable causes underlying such reversals of form.

Nor is it unreasonable to assume that those factors which operate in "persuading" a teacher to "award" to a certain individual an "A" instead of a "B," or a "B" instead of a "C," will later operate to boost the supervisor's estimate of the same individual. The two situations are enough alike, have sufficient "common elements," to suspect transfer,—more in some cases, less in others. At the same time, there are sufficient independent elements to insure against any expectation of perfect, or complete, transfer. The correlations obtained lend support to this view.

It is no longer seriously doubted that for people generally achievement in school promises success outside of school[1] (success in school and outside + .74), assuming relatively equal application of abilities and effort; and the relationship is marked even in the absence of such assumption, or equal application. What makes for success in the one makes for it in the other to a significant degree. Doubtless such traits as industry, intelligence, tact, and the like, condition whatever results are wrought. Ability and "push," other things equal, lead to success whether in school or out.

---

[1] Chassell, L. M. *The Relation Between Success in a Professional Institution and Success in a Professional Field,* Table II. [Unpublished Teachers College thesis.]

Assuming reasonable validity for the measurements made, the eight traits included under the head of "Personality" are by no means a perfect index of the success of a student or a teacher, although they possess much in common with both as indicated by the correlations. The relationship, however, does not appear to be excessive when it is considered that the educational practitioner has given high place[1] to these qualities in the successful teacher's equipment. If interests and abilities be relatively stable and permanent,[2] it is not unreasonable to expect their presence in measureable amounts at least two years in advance of the actual teaching performance,—upon entrance to the training course. Could all traits and functions related to success in teaching be identified and exact measurements of each one be secured, the resulting personality would correlate perfectly with the achievement. Were it possible in like manner to identify all qualities involved in procuring school marks and measure them with mathematical precision, the resulting correlation would be perfect. But such perfection in judgment and measurement is at present impossible and will doubtless so continue,—but with progressively diminishing amounts.

What we have, as revealed by the data, may be represented roughly by the overlapping of circles, as illustrated below. The numbers corresponding to the obtained correlations (simple, partial, and multiple) indicate the "common," "independent," and "associated" elements and the degree or amount of each. In no two cases are the circles of abilities identical,—and there is probably less identity between them in real life than the drawing indicates.

---

[1] Bagley, W. C. *School Discipline,* pp. 30-46.

[2] Thorndike, E. L. In *School and Society,* February 10, 1917.

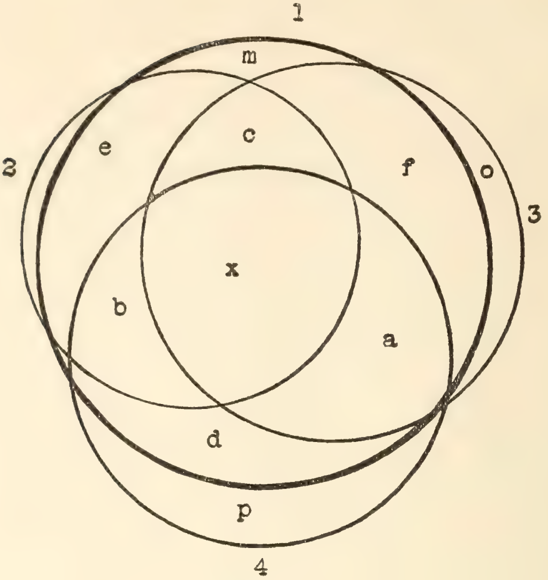


FIG. 8.

1. Circle of abilities (or functions) involved in Teaching.
2. “ “ “ “ “ involved in Reactions to Tests.
3. “ “ “ “ “ included under Personality.
4. “ “ “ “ “ involved in First Semester's Marks.

x = Those “abilities” common to the (all) four groups of activities.

bcex = “abilities” common to Teaching and Tests............................r_{12}=.425
acfx = “ “ “ “ “ Personality.........................r_{13}=.615
adbx = “ “ “ “ “ First Sem. Marks...................r_{14}=.600
cx = “ “ “ “ “ Tests and Personality.................r_{23}=.517
bx = “ “ “ “ “ Tests and First Semester Marks......r_{24}=.517
ax = “ “ “ “ “ Personality and First Sem. Marks.....r_{34}=.720

be = “abilities” common to Teaching and Tests (Pers. constant)................$r_{12\cdot 3}$=.059
ce = “ “ “ “ “ “ (1st S. M. constant)............$r_{12\cdot 4}$=.066
af = “ “ “ “ “ Pers. (Tests constant)................$r_{13\cdot 2}$=.510
cf = “ “ “ “ “ “ (1st S. M. constant)............$r_{13\cdot 4}$=.330
ad = “ “ “ “ “ 1st S. M. (Tests constant)............$r_{14\cdot 2}$=.472
bd = “ “ “ “ “ 1st S. M. (Pers. constant)............$r_{14\cdot 3}$=.288

e = “abilities” common to Teaching and Tests (Pers. & 1st S. M. constant)...$r_{12\cdot 34}$=.034
f = “ “ “ “ “ Pers. (Tests & 1st S. M. constant)....$r_{13\cdot 24}$=.325
d = “ “ “ “ “ 1st S. M. (T. & P. constant).........$r_{14\cdot 23}$=.247

abecfx = “abilities” common to Teaching and Tests and Personality............$R_{1\cdot 23}$=.628
adbecx = “ “ “ “ “ Tests and First Semester Marks....$R_{1\cdot 24}$=.603
adbcfx = “ “ “ “ “ Personality and First Sem. Marks.$R_{1\cdot 34}$=.655

adbecfx = “abilities” common to Teaching and Tests-Pers. and 1st S. M.......$R_{1\cdot 234}$=.658

m, n, o, p, the “independent” functions under 1, 2, 3 and 4, respectively.

NOTE: No attempt is made to approximate proportional areas in the drawing to represent the various correlations, which are themselves trigonometric functions rather than geometric.

THE GOODNESS OF PREDICTION

For the problem of prognosis, for predicting outcomes, the mere discovery and determination of relationships,—apparent and real, simple, partial, and multiple,—will not suffice. To estimate a person's achievement in school and outside from test scores, personality estimates, and early school marks, a method for expressing the relationship between "two concurrently changing quantities" is necessary. We need to know not only to what degree the second is connected, or associated, with the first, but the most probable value of the former on the basis of the latter. The magnitude of the correlation coefficient indicates the status of relationship to the extent of denoting whether or not position above (or below) the mean in the one promises position in the same direction in the other, —and the degree of such probability. But it does not express the most probable value of the one with a given value of the other,— the essential requirements, the *sine qua non,* of the present problem for predicting a person's standing as student in school and as teacher in the class-room outside on the basis of known test-ability, personality, and early school marks, singly and in combination.

An equation, therefore, is required which will enable us to put two or more variable quantities, as X_1 and X_2, or X_1, X_2, and X_3, in an algebraic expression in such a way that a given change in the value of one, say X_1, is accompanied by a proportional change in the value of the other, X_2; or a given change in X_1 is attended by proportional changes in X_2 and X_3; and similarly for any larger number of factors. For example, if Test Scores, Personality Estimates, and First-Semester-Marks in the original data are represented by X_1, X_2, and X_3, respectively, and if the Success in the Normal School and Teaching-in-Field, the first year, are represented by X_N and X_F, the problem is to determine the relationship between X_N (and X_F) and any one, any two, or all three of these measures. If we take the combined measures based upon X_1, X_2, and X_3, and express as an equation, it is X_N (or X_F) $= C_0 + C_1 X_1 + C_2 X_2 + C_3 X_3$.[1] That is, the most probable *success value* is equivalent to a certain-constant, plus a certain-constant-times-score-earned-in-tests, plus a certain-constant-times-personality-estimate, plus a certain-constant-times-first-semester-marks. The statistical problem involved is the determination of the constants C_0, C_1, C_2, C_3, so that the X_N *values obtained* differ on the whole, and, when every individual is taken into account, from the *actual* X_N values by the smallest amount possible.

---

[1] Kelley, T. L. *Educational Guidance,* p. 5.

The Regression Equation. The equation which fulfills this condition is called a regression equation, and the constants (other than C_0), C_1, C_2, C_3, are called regression coefficients. They are functions of the coefficients of correlation between the various X's and the standard deviation of the X's. But since the measures herein are expressed in terms of variability, the standard deviations being equal (10-tenths sigma), the regression coefficient when but two variables are involved is identical with the coefficient of correlation. The problem is "to determine the coefficients and constant term, if any, in a regression equation, so as to make the sum of the squares of the errors of estimate a minimum".[1] That is, such that when the *most-probable-success-value* has been calculated by the formula and the difference between that and the *actual-success-value* computed and these differences squared and summed producing a smaller amount than is possible under any other arrangement, the equation is the one desired, the equation of best fit. The mathematically minded will experience no difficulty in following the statistical process involved; the non-mathematically minded will have to accept it largely on faith.

The essential point to understand and to hold in mind is that the X_N value for each individual so obtained by the equation is the most probable value that can be obtained from the data X_1, X_2, X_3. The method of combining the several factors or sets of measures is that of weighting according to statistically determined values,—the different measures being given not equal weight when combined but each its own earned increment, that weight which it added to the total when considered in relation to the "associated" set (or sets) of measures and to the achievement itself. This type of equation is the means employed in the following pages for predicting success, and to apply it in the case of any individual it is only necessary to substitute the values X_1, X_2 X_3 (one, two or all three of them, depending upon the particular equation), for that particular individual to obtain her most probable X_N value. To make concrete the method and to facilitate the arithmetic, a double-entry table has been added (Table XIX, p. 78). Demands of economy have prevented inclusion of tables for more than three variables.

---

[1] Yule, G. U. *An Introduction to the Theory of Statistics,* pp. 231 ff.

The coefficients of correlation already reviewed revealed the relative closeness of relation between the different factors, or sets of measures, and the success achieved, and at the same time indicated the predictive values. The regression equations which follow will serve to emphasize those differences and show in a practical form their use as prognosis measures.

PREDICTION FROM ONE SET OF MEASURES

Equations for Two Variables

A comparison may be drawn from the following sets of equations as to the goodness of prediction of each set of measures singly for each type of achievement, and of each set of measures for the different types of achievement. In each group equation, (1) denotes Tests, (2) Personality, (3) First-Semester-Marks.

With Reference to Two-Years'-Marks. The equations which fit the data here are as follows:

$$(1) \quad X_M = .578X_1 + 21.1$$
$$(2) \quad X_M = .762X_2 + 11.9$$
$$(3) \quad X_M = .776X_3 + 11.2$$

It is well to remember that the "scores" or values are expressed in terms of variability, as tenth-sigma-measures, and considered as so-far-above or so-far-below the mean, it being at the point designated as 50-units. Equation (1) means that for an individual with a given position in Tests, say 65, her most probable position in Two-Years'-Marks will be 65 times .578 plus 21.1 which is approximately 58.7. Equation (2) yields for the equivalent position in Personality 61.4 in Two-Years'-Marks. And equation (3) gives 61.6. Similarly, for all other positions or values. Once the equation is determined, it becomes a mere matter of arithmetic to make the substitution.

The interpretation is not difficult. Equation (1) which is based upon the lowest correlation gives the least good prediction. To be up-to-standard in Test-ability does not indicate being up-to-standard in school grades to the degree that it would be "promised" by equivalent position in Personality or First-Semester-Marks. The difference in the prognostic value of the last two is negligible. Perfect prediction would yield 65 as the most probable value and without error.

With Reference to Teaching-in-Training-School. The equations which fit the data here are as follows:

$$\begin{aligned}
&(1)\quad X_T = .484X_1 + 25.8\\
&(2)\quad X_T = .608X_2 + 19.6\\
&(3)\quad X_T = .632X_3 + 18.4
\end{aligned}$$

The order here is consistent with the relative value of the three measures for predicting success in *practice teaching.* The "best," or most probable, values for the training school work on the basis of a 65-position in Tests, Personality, and First-Semester-Marks, are 57.3, 59.1, and 60.5, respectively. The prediction in each case is less close than for school marks as shown above, due to lower correlations.

With Reference to Normal-School-Success. The equations which best fit the data for the whole work in the normal school in relation to the three sets of measures are:

$$\begin{aligned}
&(1)\quad X_N = .540X_1 + 23.0\\
&(2)\quad X_N = .720X_2 + 14.0\\
&(3)\quad X_N = .724X_3 + 13.8
\end{aligned}$$

Again, the order of prognostic value increases with the series of equations,—very much from (1) to (2) and only slightly from (2) to (3). The "best" values for the whole of the normal school work, given by these equations on the basis of a 65-position in each of the three measures, are 58.1, 60.8, and 60.9, respectively.

With Reference to Teaching-in-the-Field. The equations which best fit the data here are as follows:

$$\begin{aligned}
&(1)\quad X_F = .425X_1 + 28.75\\
&(2)\quad X_F = .615X_2 + 19.25\\
&(3)\quad X_F = .600X_3 + 20.00
\end{aligned}$$

The order is changed this time in one particular, as suggested by the size of the constant factor. Personality, which yields a correlation slightly higher with Teaching than do the First-Semester's-Marks, gives a slightly closer prediction of the level of success to be achieved. The most probable values given by these equations for an individual's level of achievement in Teaching on the basis of a 65-position in Tests, Personality, and First-Semester-Marks are 56.4, 59.2, and 59, respectively.

PREDICTION FROM A COMBINATION OF TWO SETS OF MEASURES

Equations for Three Variables

Comparisons may be made of the equations to follow as to the goodness of prediction of each combination of measures for the several sorts of success. Equation (1) signifies Tests and Personality; (2) Tests and First-Semester-Marks; (3) Personality and First-Semester-Marks. The regression equations giving that combination of these factors taken two-at-a-time which correlates the highest with, and which yields the best predictions for, the particular type of achievement are:

With Reference to Two-Years'-Marks. The equations which fit the data here are as follows:

$$(1)\quad X_M = .25X_1 + .63X_2 + 6.0$$
$$(2)\quad X_M = .138X_1 + .668X_3 + 9.7$$
$$(3)\quad X_M = .423X_2 + .472X_3 + 5.25$$

Equation (1) gives a somewhat closer prediction than equation (2) and only slightly less close than (3). The most probable success values in Two-Years'-Marks yielded by these equations for the individual whose position is at level 65, or its equivalent (see Table XIX, p. 78) in each of the measures in the combination are 63.2, 62.1, and 63.4, respectively. Perfect prediction would give 65 as the "best" value with negligible error.

With Reference to Teaching-in-Training-School. The equations which fit the data here are:

$$(1)\quad X_T = .23X_1 + .49X_2 + 14.0$$
$$(2)\quad X_T = .134X_1 + .546X_3 + 16.0$$
$$(3)\quad X_T = .318X_2 + .400X_3 + 14.1$$

Equation (1), as suggested by the smallness of the constant term, gives the best prediction, much better than (2) and slightly closer than (3). The respective "best" values for position 65 are 60.8, 60.2, and 60.8. Tests in the first combination seem to contribute relatively much more (weighted 1 : 2) than in the second (1 : 4), due largely to its lower correlation with Personality than with First-Semester-Marks, the variability in the two cases being not appreciably different. The reduced weights for Personality and First-Semester-Marks in combination are due in large part to the overlapping of functions as related to practice teaching,—each accounting for much in the other.

With Reference to Normal-School-Success. The equations which fit the data here are:

$$
\begin{aligned}
&(1)\quad X_N = .229X_1 + .602X_2 + 8.45\\
&(2)\quad X_N = .130X_1 + .649X_3 + 11.05\\
&(3)\quad X_N = .412X_2 + .432X_3 + 7.80
\end{aligned}
$$

A combination of *Personality and First-Semester-Marks* here gives a slightly better prediction of the complete normal school performance of an individual than that yielded by *Tests and Personality* combined, either of these being somewhat better than *Tests and First-Semester-Marks* in combination. The most probable values in Normal-School-Success given by these equations for the 65-position in each set of measures (or its equivalent. See Table XIX below) in the combination are 62.4, 61.7, and 62.7.

With Reference to Teaching-in-the-Field, First Year. The equations which fit the data here are:

$$
\begin{aligned}
&(1)\quad X_F = .146X_1 + .540X_2 + 15.7\\
&(2)\quad X_F = .069X_1 + .555X_3 + 18.8\\
&(3)\quad X_F = .380X_2 + .326X_3 + 14.7
\end{aligned}
$$

The order of value of these combinations, and of the equations, for prognostic purposes is identical with what was noted in the case of the immediately preceding set of equations. These are, however, much less close in every instance than those above,—the best one here (3) being less pointed and exact in its prediction than the poorest one (2) there,—although not appreciably poorer than those determined for Teaching-in-the-Training-School from the same combinations. The "best" values given by these equations for the 65-position are 60.3, 59.4, and 60.6, respectively.

Comparing the predictions possible on the basis of a combination of measures (factors taken two-at-a-time) with those from a single set of measures (factors taken singly), the superiority of the former is clear. The correlation is higher generally, and in many cases appreciably so, and the prediction is closer with a correspondingly reduced error. Table XX below (p. 82) which summarizes the principal regression equations, correlation coefficients between the theoretical and actual values (i.e., between those values predicted by the equation and the values found to exist from the actual data), and the variabilities (S.D.'s and P.E.'s), will serve to make the differences and superiorities clear.

TABLE XIX

Regression Values for Normal-School-Success on $\left.\begin{matrix}\text{MENTAL TESTS}\\ \text{PERSONALITY}\end{matrix}\right\}$

$$(X_N = .229X_1 + .602X_2 + 8.45)$$

POSITIONS (TENTH-SIGMA) IN TESTS	POSITIONS (TENTH SIGMA MEASURES) IN PERSONALITY																	
	71	68	65	62	59	56	53	50	47	44	41	38	35	32	29	26	23	20
71	*67.4*	65.6	63.8	62.0	60.2	58.4	56.6	54.8	53.0	51.2	49.4	47.6	45.8	44.0	42.2	40.4	38.6	36.8
68	66.8	*65.0*	63.2	61.4	59.6	57.8	56.0	54.2	52.4	50.6	48.8	47.0	45.2	43.4	41.6	39.8	38.0	36.2
65	66.1	64.3	*62.5*	60.7	58.9	57.1	55.3	53.5	51.7	49.9	48.1	46.3	44.5	42.7	40.9	39.1	37.3	35.5
62	65.4	63.6	61.8	*60.0*	58.2	56.4	54.6	52.8	51.0	49.2	47.4	45.6	43.8	42.0	40.2	38.4	36.6	34.8
59	64.7	62.9	61.1	59.3	*57.5*	55.7	53.9	52.1	50.3	48.5	46.7	44.9	43.1	41.3	39.5	37.7	35.9	34.1
56	64.0	62.2	60.4	58.6	56.8	*55.0*	53.2	51.4	49.6	47.8	46.0	44.2	42.4	40.6	38.8	37.0	35.2	33.4
53	63.3	61.5	59.7	57.9	56.1	54.3	*52.5*	50.7	48.9	47.1	45.3	43.5	41.7	39.9	38.1	36.3	34.5	32.7
50	62.6	60.8	59.0	57.2	55.4	53.6	51.8	*50.0*	48.2	46.4	44.6	42.8	41.0	39.2	37.4	35.6	33.8	32.0
47	61.9	60.1	58.3	56.5	54.7	52.9	51.0	49.3	*47.5*	45.7	43.9	42.1	40.3	38.5	36.7	34.9	33.1	31.3
44	61.3	59.4	57.6	55.8	54.0	52.2	50.4	48.6	46.8	*45.0*	43.2	41.4	39.6	37.8	36.0	34.2	32.4	30.6
41	60.5	58.7	56.9	55.1	53.3	51.5	49.7	47.9	46.1	44.3	*42.5*	40.7	38.9	37.1	35.3	33.5	31.7	29.9
38	59.8	58.0	56.2	54.4	52.6	50.8	49.0	47.2	45.4	43.6	41.8	*40.0*	38.2	36.4	34.6	32.8	31.0	29.2
35	59.2	57.4	55.6	53.8	52.0	50.2	48.4	46.6	44.8	43.0	41.2	39.4	*37.6*	35.8	34.0	32.2	30.4	28.6
32	58.5	56.7	54.9	53.1	51.3	49.5	47.7	45.9	44.1	42.3	40.5	38.7	36.9	*35.1*	33.3	31.5	29.7	27.9
29	57.8	56.0	54.2	52.4	50.6	48.8	47.0	45.2	43.4	41.6	39.8	38.0	36.2	34.4	*32.6*	30.8	29.0	27.2
26	57.1	55.3	53.5	51.7	49.9	48.1	46.3	44.5	42.7	40.9	39.1	37.3	35.5	33.7	31.9	*30.1*	28.3	26.5
23	56.4	54.6	52.8	51.0	49.2	47.4	45.6	43.8	42.0	40.2	38.4	36.6	34.8	33.0	31.2	29.4	*27.6*	25.8
20	55.7	53.9	53.1	50.7	48.5	46.7	44.9	43.1	41.3	39.5	37.7	35.9	34.1	32.3	30.5	28.7	26.9	*25.1*

NOTE.—The "best value" or most probable position in Normal-School-Success for an individual whose "score" in Tests is 68 and whose "score" in Personality is 71 will be 66.8. For one whose "score" in Tests is 65 and whose "score" in Personality is 50 it will be 53.5. To obtain any "best value" from this table find first the "score" in Tests in column at left and then follow numbers in body of table from left to right until the "score" in Personality (row of figures in upper part of table) is reached, the number directly under it is the number desired.

PREDICTION FROM A COMBINATION OF THREE SETS OF MEASURES

Equations for Four Variables

To determine the total bearing of the three sets of measures upon the four types of achievement, the regression equations combining them have been derived. The equation which best fits the data for Tests, Personality, and First-Semester-Marks, in combination, is:

With Reference to Two-Years'-Marks:

$$X_M = .106X_1 + .412X_2 + .420X_3 + 3.1$$

This equation gives a somewhat better prediction than that obtained from any single measure or from a combination of any two. The most probable value here for our familiar position, 65, is 64, approximately,—Personality and First-Semester-Marks each contributing about four times as much toward the prognosis as do Tests. $R_{M \cdot 123} = + .833$, with a probable error of $\pm .017$, which means that, given an individual's record, or "score," in these three types of performance, we can predict with a relatively high degree of certainty and with small error the level of achievement over the subsequent year and a half in the normal school,—a somewhat closer forecast than is possible on the basis of grades alone or of teachers' estimates.

With Reference to Teaching-in-Training-School:

$$X_T = .105X_1 + .306X_2 + .345X_3 + 12.2$$

The prediction possible from this equation is not considerably superior to what Tests and Personality would give freed from First-Semester-Marks. The "best" value for the 65 position from this is only slightly above 61.3,—the two giving 60.8. All other things considered, it is by no means clear that the institution would be justified in holding the recruit in school for the half year for the additional value this record brings to the guidance problem. If later data show the suggestion to be well founded, the wisdom of early action in the matter would be clear; and should the present data for the complete normal school record and for the achievement in teaching outside also give evidence of similar small additions from First-Semester-Marks, it will be apparent that the facts here are by no means accidental.

With Reference to Normal-School-Success:

$$X_N = .093X_1 + .406X_2 + .379X_3 + 6.1$$

The prediction here possible is not a great deal closer than from *Tests and Personality* combined. The "best" value given for position 65 is a little less than 63.2, against 62.4 on the basis of those two. The difference between the values of the *three-combined* and the *two-combined* is far less than that between the latter and any single set of measures. The amount added by First-Semester-Marks is neither great nor negligible as indicated by the increased correlation between theoretical and actual values (from .746 to .781), the reduction of the S.D. (from 6.66 to 6.24), and the better predicted values (as above, from 62.4 to 63.2). The data on this point are far from being conclusive,—clearly inadequate to a definite conclusion as to the justification for a probationary period for the half year.

There can be little doubt of the added value which the record accumulated the first half year gives to the prediction of final school achievement. It is a question, however, as to whether or not the prognostic value added justifies retaining recruits for orientation purposes, four months or more,—whether the increment of justice to the few who might on the basis of Tests and Personality be "advised out," compensates for the injustice to possibly many more who would be required to remain for several months to learn of their unsuitability for the teaching profession. In the light of facts so far presented, exclusion on the basis of *Tests and Personality* alone would not seem defensible, in so far as achievement in the normal school as a whole is concerned. Further consideration will be given to this point in Chapter VI, and in a more tangible and concrete form.

With Reference to Teaching-in-Field. Two regression equations have been derived for predicting success in teaching,—(1) From data available at the end of the first half year, Tests, Personality, and First-Semester-Marks; (2) from data available at the end of the two years' training course, Tests, Personality, and Normal School records.

$$(1) \quad X_F = .033X_1 + .376X_2 + .311X_3 + 14.00$$

$$(2) \quad X_F = .009X_1 + .177X_2 + .599X_N + 10.75$$

In equation (1), relatively less weight is given to Test-ability than that given it in the corresponding equation for "practice teaching." At the same time, greater weight is given to Personality in this than it received in that, while First-Semester-Marks get approximately the same in both. This fact seems to indicate that ability as measured by tests has much less weight in its effect on the supervisor's judgment than has the general personality itself, especially in the field. With school records and personality equal, abstract intelligence would appear to be not a large factor in winning the approval of one's *superior officers;* but with only Personality equal, the weight of Test-ability increases approximately 300 per cent.

Again, the value added by the First-Semester-Marks is relatively about the same as noted above in connection with the complete normal school record. It is neither negligible nor considerable in amount. Equation (1) here with First-Semester-Marks included in the combination is about midway in its prognostic directness and precision between the corresponding equation for *Tests and Personality* combined, and equation (2) in the present connection for *Tests, Personality, and Normal-School-Success.*

The complete normal school record as indicated by equation (2) as against (1) above and equation (1) for *Tests and Personality* (p. 76) shows appreciable relation to and general value for predicting the level of success in Teaching. The prognostic value of the three combined, for Teaching, is practically identical with that of *Tests and Personality* alone for achievement in school. The correlation between "values by formula" and "actual values" for the two combination are +.748 and +.746, respectively, with S.D.'s of 6.64 and 6.66; and values for the now well-known 65 position of 61.8 and 62.4.

Little doubt remains as to the value added by the record made in school, but the question as to whether or not we are justified in subjecting students to two full years of training in order to arrive at a basis for reasonable guidance, whether or not serious injustice to the cause would result from practical guidance given earlier in the course, or even at the beginning, on the basis of Test-ability, Personality and possibly high school records, is a matter yet to be determined.

The principal equations derived from the data are summarized as follows:

TABLE XX

SIGNIFICANT REGRESSION EQUATIONS FOR PREDICTING ACHIEVEMENT

1. EQUATIONS FOR THREE VARIABLES

					*	**
With reference to		S.D.	*R*	P. E.	Measure	Predicted
Two-Years'-Marks:	(1) $X_M = .25X_1 + .63X_2 + 6$	6.11	.792	±.020	65	63.2 ±.61
	(2) $X_M = .138X_1 + .668X_3 + 9.7$	6.22	.783	±.021	65	62.1 ±.62
	(3) $X_M = .423X_2 + .472X_3 + 5.25$	5.59	.828	±.017	65	63.4 ±.56
Teaching-in-Training-School	(4) $X_T = .23X_1 + .49X_2 + 14$	7.68	.640	±.032	65	60.8 ±.76
	(5) $X_T = .134X_1 + .546X_3 + 16$	7.68	.640	±.032	65	60.2 ±.76
	(6) $X_T = .318X_2 + .400X_3 + 14.1$	7.39	.676	±.029	65	60.8 ±.73
Normal-School-Success	(7) $X_N = .229X_1 + .602X_2 + 8.45$	6.66	.746	±.024	65	62.4 ±.66
	(8) $X_N = .130X_1 + .649X_3 + 11.05$	6.83	.732	±.025	65	61.7 ±.68
	(9) $X_N = .412X_2 + .432X_3 + 7.80$	6.28	.779	±.021	65	62,7 ±.62
Teaching-in-the-Field	(10) $X_F = .146X_1 + .540X_2 + 15.7$	7.78	.628	±.039	65	60.3 ±.89
	(11) $X_F = .069X_1 + .555X_3 + 18.8$	7.98	.603	±.041	65	59.4 ±.91
	(12) $X_F = .380X_2 + .326X_3 + 14.7$	7.55	.655	±.036	65	60.6 ±.86

2. EQUATIONS FOR FOUR VARIABLES

					*	**
With Reference to		S.D.	*R*	P. E.	Measure	Predicted
Two-Years'-Marks	(1) $X_M = .106X_1 + .412X_2 + 420X_3 + 3.1$	5.53	.833	±.016	65	64.0 ±.55
Teaching-in-Training School	(2) $X_T = .105X_1 + .306X_2 + .345X_3 + 12.2$	7.37	.680	±.029	65	61.3 ±.74
Normal--School-Success	(3) $X_N = .093X_1 + .406X_2 + .379X_3 + 6.1$	6.24	.781	±.020	65	63.2 ±.62
Teaching-in-the-Field	(4) $X_F = .033X_1 + .376X_2 + .311X_3 + 14.0$	7.53	.658	±.036	65	61.1 ±.85
	(5) $X_F = .009X_1 + .177X_2 + .599X_N + 10.75$	6.65	.748	±.031	65	61.8 ±.76

*The number 65 here is made constant throughout as values for X_1, X_2, X_3 simply for purposes of illustration and comparison. Any other value (or values) might be used.

**The numbers predicted are the averages of the arrays (which are themselves the best single values) plus or minus so much. The probable error in each case refers to the particular average. See *Biometrika*, Vol. IX, p. 10.

Two sets of facts stand out from the comparison of these equations. In the first place, the equations for four variables, in which achievement is predicted on the basis of the three sets of measures, prove in every corresponding case superior to those for three variables, in which prediction is based upon a combination of only two factors. The resulting correlation (multiple *R*) is higher, the variability (S.D.) reduced, and the "best" values more nearly approach the basis values as a limit, and with smaller probable error. For example, see the first equation in the second part, as compared with any one of the first three equations in the first part of the table, or the second one in the second part with any one of the next three in the first part.

In the second place, school marks are predicted by means of Tests, Personality, and First-Semester-Marks, singly or in combination, more easily and with greater certainty than is "practice teaching," and the complete school record more closely than actual teaching in the class-room outside. There appear higher correlations, less variability, closer approximations of predicted values for the former, with less error; inversely for the latter.

High School Records and Achievement. Regression equations were determined from the data for the twenty-three (23) individuals who were prepared in the high school department already referred to (p. 54). The achievement in the normal school could be prophesied with a high degree of certainty on the basis of combinations of School Marks, Personality estimates, and Tests, as indicated in the following:

X_{Ns}	X_1	X_2	X_3	Constant	S.D.	*R*	P. E.	Basis Value	Predicted Value	P.E.
(1)$X_{Ns}=$	$.622X_{HM}$	$+.284X_{HP}$		+ 4.7	5.50	.835	± .041	65	63.6	±1.9
(2)$X_{Ns}=$	$.537X_{HM}$	$+.213X_{T}$		+12.5	5.64	.826	± .042	65	61.3	±2.0
(3)$X_{Ns}=$	$.394X_{1st\,s}$	$+.544X_{NP}$		+ 3.1	4.37	.899	± .026	65	64.1	±1.6
(4)$X_{Ns}=$	$.543X_{HM}$	$+.245X_{HP}$	$+.160X_{T}$	+ 2.6	5.35	.845	± .039	65	64.2	±1.9
(5)$X_{Ns}=$	$.221X_{1st\,s}$	$+.598X_{NP}$	$+.163X_{T}$	+ 0.9	4.23	.906	± .024	65	64.7	±1.5

Personality estimates given by the normal school instructors for this small group are closer to later achievement than those by high school teachers, and First-Semester-Marks are likewise closer than high school grades. Consequently, prediction is better on the basis of the former than on the basis of the latter. That is, equation (3) above, based upon First-Semester-Marks (*1st S*) and Personality estimates given at the end of the first half year by normal school instructors (*NP*), yields higher correlation and less variability than equation (1), based upon high school marks (*HM*) and Personality estimates by the high school staff (*HP*), and the former gives a closer prophecy. Similarly for (5) as compared with (4); but in neither case is the difference great, although appreciable in both cases.

The correlations for this group as compared with the whole group (156) are somewhat high, and the equations are appreciably better in general. See (3) above as compared with (9) in the first part of Table XX, and (5) here with (3) in the second part of the same table. While it indicates that this is a select group, in that it is not 100 per cent typical of the whole group, the differences are not so great as to render them incomparable for our purposes. The smallness of the group makes impossible sweeping conclusions on the basis of the figures; but after making every allowance for the increase due to the common factor of the school situation and the continued sameness of atmosphere, the chance occurrence of a few at the extremes, low and high, there is still a residual relationship between the measures and achievement, which is extremely suggestive and which can be accounted for only on the basis of reality.

Granting the improbability (even impossibility) of procuring records (grades and trait estimates) from a large number of high schools (some large, many small,—well-staffed, and poorly-staffed) which will yield so high correlations, or even nearly so high (possibly two or three tenths lower on the average) and making all necessary discounts, this source still seems rich in possibilities for the guidance problem. If a uniform system could be inaugurated for recording not only the school grades but also the teachers' judgments on certain traits,—and the practicability is hardly to be questioned as records of some sort are universally kept,—it would supply data impossible to evaluate at present but rich with promise for practical guidance and direction of the high school graduate.

If human character be reasonably permanent,[1] if the qualities of the man and the woman be present in the boy and the girl, and evidence is accumulating in support of the assumption; and if a certain modicum of personality traits be a *sine qua non* to success (as success is measured at present) in the school room and outside (and school administrators and supervisors are in general agreement in placing a premium on certain traits and qualities in the teachers as contributing factors to success); the one outstanding problem in educational and vocational guidance for teachers is to identify such traits, measure them, and give them direction toward the end which will result in the least waste and maximum returns. If *students* are capable of making valid judgments on character traits of their fellows[1] and on teaching ability of their instructors,[2] the teachers themselves might well be expected to render even more reliable judgments on the qualities of students well-known to them; and, if it be possible to measure and evaluate them in intelligible units of amount, or by levels, whether by personal estimate or by means of objective tests, the problem of teacher guidance will be brought one step nearer to practical solution. The present study is an attempt to measure the degree to which certain traits are present in the recruit and to determine the amount of correspondence between them and later achievement,—with the results herein enumerated.

---

[1] Thorndike, E. L. In *School and Society,* Feb. 10, 1917.

COMPARISON OF LEVELS IN PER CENTS

Single Measures and Achievement. Tables XXI and XXII pp. 87-90 show the amount of correspondence discovered between positions in Tests, Personality, and First-Semester-Marks, singly and in combination, and the levels of achievement in school and teaching outside, for the 156 and 110 cases, respectively. Students were first distributed on the basis of their standing in the three sets of measures or combinations, and then redistributed (or placed") among the corresponding levels in achievement. The diagonal running through the corresponding compartments (represented by numbers in italics) cuts through the high, low, and intermediate paired groups and serves as a line of reference. The greater the concentration about this line, and the less the spread, the better the particular measure for purposes of placement within these limits. If all numbers were concentrated within these compartments,—100 in each case, expressed in per cents,—the correspondence would be perfect. The larger the numbers, therefore, at this point,—the nearer to 100,—the range of spread being constant, the better the measure; and the less the spread in extent, in its range, with the numbers about the diagonal being the same, the better the measure.

---

[1] Hollingworth, H. L. *Vocational Psychology,* Chapter VII.

[2] Knight, B. F., *op. cit.,* pp. 35-36.

As a numerical representation of the goodness-of-placement, the average percentage deviation (last column on the right) has been computed for each set of measures for each segment and for all five segments combined. Perfect placement within the limits would be represented by 0; less than perfect by a number between 0 and 400 for the "A" and "E" groups, between 0 and 300 for the "B" and "D" groups, between 0 and 200 for the "C" group; and for the general average, between 0 and 280,—each segment being weighted according to the percentage of individuals represented in it. In this case, 280 would indicate "perfect" misplacement,—those individuals who rank high in Tests (Personality, or First-Semester-Marks) would be placed low in achievement, and those ranking low would be placed high, and those just average at the two extremes, lowest and highest. Zero would obtain in the event of plus one (or nearly that) correlation, and 280 if the correlation were minus one (or nearly so). Placement by pure chance with zero correlation would result roughly in an even distribution throughout with an average percentage deviation of approximately 144.

It is apparent from both the overlapping of groups (three in some cases; in the rest, four) with the consequent reduction in size of the numbers about the diagonal and the magnitude of the placement errors (last column on the right) that the correspondence, the placement, is in no case perfect, nor nearly so. On the other hand, it is far different from what we should expect from chance. Relationship is evident between each of the measures and achievement, both in school and in the field, but not equal.

The numbers along the horizontals for Tests are in general smaller (and the spread heavier) than for Personality and First-Semester-Marks, thus confirming the facts of Tables XV, XVI and XVII, and the average percentage misplacement is also in agreement. These being matters of relative values, they do not imply non-worth of Tests.

TABLE XXI

COMPARISON OF POSITIONS IN TESTS, PERSONALITY, AND FIRST-SEMESTER-MARKS WITH LEVELS OF ACHIEVEMENT

1. IN NORMAL SCHOOL

	*See below.	*Upper 10%* "A"	*Next 20%* "B"	*Middle 40%* "C"	*Next 20%* "D"	*Lower 10%* "E"	*% Error*
"A"	Tests	*47*	20	33			.86
	Personality	*67*	13	20			.53
	1st-Sem.-Marks	*60*	33	7			.47
"B"	Tests	18	*30*	36	6		.66
	Personality	19	*47*	31	3		.56
	1st-Sem.-Marks	13	*42*	45			.58
"C"	Tests	3	23	*44*	24	6	.65
	Personality	2	18	*50*	24	6	.58
	1st-Sem. -Marks	2	19	*55*	18	6	.53
"D"	Tests	3	9	37	*28*	23	.87
	Personality		3	52	*36*	9	.67
	1st-Sem.-Marks		3	31	*53*	13	.50
"E"	Tests			27	53	*20*	1.07
	Personality			13	40	*47*	.66
	1st-Sem.-Marks			13	47	*40*	.73

Weighted Average .76 / .60 / .55

2. IN THE FIELD, FIRST YEAR

"A"	Tests	*27*	27	37	9		1.28
	Personality	*55*	9	27	9		.90
	1st-Sem.-Marks	*55*	27	18			.63
"B "	Tests	25	*33*	25	9	8	.92
	Personality	17	*44*	30	9		.65
	1st-Sem.-Marks	10	*29*	47	14		.85
"C"	Tests	8	15	*55*	17	5	.58
	Personality	3	18	*51*	23	5	.57
	1st-Sem.-Marks	8	18	*49*	15	10	.69
"D"	Tests		12	42	*25*	21	.87
	Personality		15	56	*18*	11	.96
	1st-Sem.-Marks		19	38	*35*	8	.84
"E"	Tests			45	37	*18*	1.27
	Personality			18	27	*55*	.63
	1st-Sem.-Marks			27	27	*46*	.81

Weighted Average .85 / .71 / .76

*Expressed in per cents and should read—Of those in the highest 10 per cent ("A" group) in Tests 47 per cent achieved place among the highest 10 per cent ("A" success) in Normal School; 20 per cent achieved "B" success; and 33 per cent "C" success. Likewise for Personality and First-Semester-Marks.

Just how valuable or useful Tests are for predicting achievement will appear from the facts themselves. Thus, of those who stood in the highest 10% group (and we may refer to the groups by letters) in Tests, 47% achieved "A" success in school; 20%, "B" success; 33%, "C" success; none below. Only 6% of those in the next 20% in Tests fell as low as "D" in school; none at "E". Of the middle 40% group, only 3% achieved "A" success, and but 6% "E" success (failure). Of the next 20% (or "D" group), only 12% reached the "B" level of achievement,—3% being in the "A" group. And of those in the lowest 10% (or "E" group), none achieved success in school above the "C" level. While we cannot predict very closely the achievement, we can say what it will be within certain limits. That is, individuals who test at "A' level will achieve position in normal school work of "C" grade or better; those at "B" level will achieve "D" success or better (most cases better); a few (very few) of those at "C" level may reach as high as "A" success, and a few (very few) as low as "E"; a few of those at "D" level may achieve "A" grade, and a few more "B" grade; and those at "E" level will achieve place not above "C."

Prediction from Personality and First-Semester-Marks will prove closer on the whole for success in school and about as close for the first year's teaching in the field as in the case of Tests for in-school-achievement. This may be read from the detailed figures in the table; it is indicated by the weighted-average-percentage-misplacement for each at the extreme right, on the margin.

The facts of the case would not justify the hope of being able, on the basis of Tests alone, closely to approximate the achievement as student or teacher with correlations ranging from +.425 (with teaching) to +.578 (with two-year-marks). Nor would Personality with correlations from +.615 (with teaching) to +.762 (with two-year-marks) be a sufficient basis, in and of itself, for predicting that outcome. Nor would First-Semester Marks be sufficient as such. In fact, no one set of measures would serve adequately to predict both types of achievement, being somewhat below the correlation (plus .8 or plus .9) considered necessary for reliable prediction.[1]

But to say that a measure, or set of measures, is inadequate alone does not imply lack of value when combined with other measures, or sets of measures. In fact, it may possess great value, and frequently does, when so combined. Nor does it necessarily hold that because one measure, or set of measures, has in itself much predictive worth, it will add correspondingly great weight when combined with other measures, or sets of measures. This is conditioned on its relation to the associated measures, or sets of measures, in the combination,—if closely related to them, it adds little; conversely, much.

---

[1] Thorndike, E. L. In *School and Society,* Feb. 15, 1919.

Combinations of Measures and Achievement. In Table XXII, a comparison is made of the relative worth of the two *combinations* for purposes of predicting in-school and in-field achievement.

TABLE XXII

COMPARISON OF POSITIONS IN COMBINATIONS (REGRESSION VALUES) WITH LEVELS OF ACHIEVEMENTS

1. IN NORMAL SCHOOL

	**See Below	*Upper 10%* "A"	*Next 20%* "B"	*Middle 40%* "C"	*Next 20%* "D"	*Lower 10%* "E"	*% Error*
"A"	Tests and Personality	*54*	33	13			.59
	Tests, Personality and 1st-Sem.-Marks	*53*	40	7			.54
"B"	Tests and Personality	13	*50*	37			.50
	Tests, Personality and 1st-Sem.-Marks	22	*47*	31			.53
"C"	Tests and Personality	3	17	*58*	19	3	.48
	Tests, Personality and 1st-Sem.-Marks		16	*65*	16	3	.38
"D"	Tests and Personality		3	32	*49*	16	.54
	Tests, Personality and 1st-Sem.-Marks		3	29	*48*	20	.55
"E"	Tests and Personality			13	33	*54*	.59
	Tests, Personality and 1st-Sem.-Marks			6	47	*47*	.59

Weighted Average {.52
.48

2. In Teaching Outside

"A"	Tests and Personality	*46*	27	18	9		.90
	Tests, Pers. and 1-SM.	*58*	42				.42
	Tests, Pers. and N SS.	*64*	36				.36
"B"	Tests and Personality	23	*38*	35	4		.66
	Tests, Pers. and 1 SM.	10	*36*	45	9		.63
	Tests, Pers. and N SS.	14	*41*	45			.59
"C"	Tests and Personality	2	19	*51*	24	4	.55
	Tests, Pers. and 1 SM.	2	16	*56*	23	3	.49
	Tests, Pers. and N SS.		22	*57*	17	4	.47
"D"	Tests, and Personality		11	44	*28*	17	.83
	Tests, Pers. and 1 SM.		14	36	*32*	18	.82
	Tests, Pers. and N SS.			41	*47*	12	.53
"E"	Tests and Personality			27	27	*46*	.81
	Tests, Pers. and 1 SM.			20	30	*50*	.70
	Tests, Pers. and N SS.				45	*55*	.45

Weighted Average { .69 .60 .49

*Expressed in per cents and should read: Of those in the highest 10% ("A" group) in *Tests and Personality* 54% achieved place among the highest 10% ("A" success) in Normal School; 33% achieved "B" success; and 13% "C" success. Of those in the highest 10% ("A" group) in *Tests, Personality, and First-Semester-Marks* 53% achieved place among the highest 10% ("A" success) in Normal School; 40% achieved "B" success; and 7% "C" success. Similarly for the rest of the table.

If the facts of Tables XXI and XXII are compared, it is clear that the correspondence for the "combination" values with school and teaching success is closer than for the values from single measures with that success; and the three (Tests, Personality, and First-Semester-Marks) combined into a single measure, are better than any two in such combination. The per cents in close proximity to the diagonal are thereby increased, and the numbers distant from it, together with the spread, are reduced in proportion. The percentage misplacement errors show similar reduction.

On the whole, *Tests and Personality* seem to predict the achievement in school ($r = + .746$) somewhat more closely than do these together with First-Semester-Marks forecast success in the field ($r = + .658$). Or about as well as Tests, Personality, and the entire normal school record combined predict teaching success ($r = + .748$). Suggestions as to the advisability of eliminating recruits on the basis of such data will be considered in detail in the following chapter on the Disposition of the Extreme Cases.

CHAPTER VI

THE DISPOSITION OF THE EXTREME CASES

The Crux of the Problem of Prognosis

Although, as is highly desirable, the problem of prognosis includes a consideration of every member of the group, yet the essential pivotal value hinges on the guidance given to the few destined otherwise to achieve failure in the field. The large majority of recruits will do satisfactory work in the Normal School and will, after graduation, make an acceptable teaching record outside; they require relatively little guidance and direction.

In the preceding chapter it was seen that, by means of the regression equation, the achievement of the "average" student in school and the graduate in the field may be predicted within relatively narrow limits. Tests and Personality combined will serve to predict the former; these two in combination with First-Semester-Marks or Normal-School-Success may be used as a basis for predicting the latter.

But what becomes of the extremely *high* and extremely *low* cases in Tests, Personality, and First-Semester-Marks—these considered singly and in combination? What are their levels of achievement and in what proportion do they attain to the various levels? To what extent does relatively high or low rank in the one "promise" correspondingly high or low rank in success in the other? And to what extent would it be possible to counsel aright both those who are eliminated during the first year and those who achieve low position at the end of the two years' course?

It is the purpose of this chapter to trace and measure the correspondence between high place (highest 10 per cent) in Tests, Personality, and First-Semester-Marks—and in combinations of these—and achievement in school and in the field: also between low rank (lowest 10 per cent) in these measures and their achievement; furthermore, to find to what extent those attaining to the highest and lowest tenths are drawn from relatively high and low positions in Tests, Personality, and First-Semester-Marks and their combinations.

DESTINATION OF THE EXTREME VARIANTS

(1) *As Shown by the Levels of Achievement in Single Measure.* To the question, "To what extent do high and low scores in Tests, Personality, and First-Semester-Marks, give promise of achievement in two years' grades, teaching in training school, normal school success, and in-field-success, first year?"—Plates 3, 4, 5, 6 below indicate the levels of achievement together with the per cents attaining to each level in each of these four fields for the "highest" and "lowest" tenths in Tests, Personality, and First-Semester-Marks, respectively.

For example (Fig. 9, Plate 3,) of the highest ten per cent (15 individual in Tests, 40% (or 6) of them were found among the highest ten per cent (or "A" group) in the two years' grades; 33.3% (or 5) among the next twenty per cent (or "B" group); and 26.7% (or 4) among the middle forty per cent (or "C" group). Of the lowest ten per cent (15 cases) in Tests, 26.7% (or 4) of them were numbered among the lowest ten per cent (or "E" group) in two years' grades; 40% (or 6) among the next twenty per cent (or "D" group); and 33.3% (or 5) among the middle forty per cent (or "C" group); with one case placed slightly above the average for the group as a whole.

A like interpretation is to be applied to the other graphs in the series, the heading indicating to what it refers.

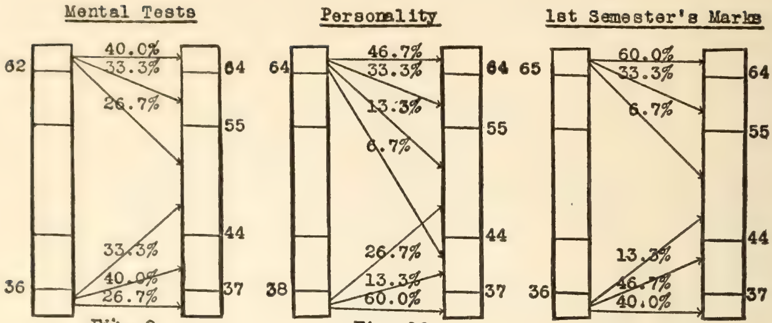


Fig. 9 Fig. 10 Fig. 11

Plate 3. Levels of Achievement in Two-Years'-Marks of Highest 10 Per Cent and Lowest 10 Per Cent.

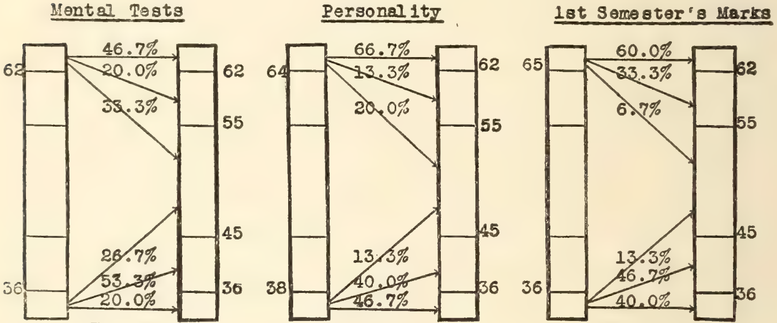


Fig 12 Fig. 13 Fig 14

Plate 4. Levels of Achievement in Normal-School-Success of Highest 10 Per Cent and Lowest 10 Per Cent.

Note: The numbers at the divisions indicate the levels in the different measures and those on the connecting lines the per cents of individuals coming from or going to each of the five groups. For example, 62 at upper left (Fig. 9) above marks the lower limit of the highest 10 per cent in Tests, the 36 directly under, marks the upper limit of the lowest 10 per cent in Tests. The 64 near the top of the right half marks the lower limit of the highest 10 per cent in Two-Years'-Marks; 64 and 55 mark the upper and lower limits respectively of the next 20 per cent in Two-Years'-Marks; 55 and 44 the upper and lower limits of the middle 40 per cent; 44 and 37 the upper and lower limits of the next 20 per cent; and 37 the upper limit of the lowest 10 per cent. A like interpretation of the numbers in all similar figures.

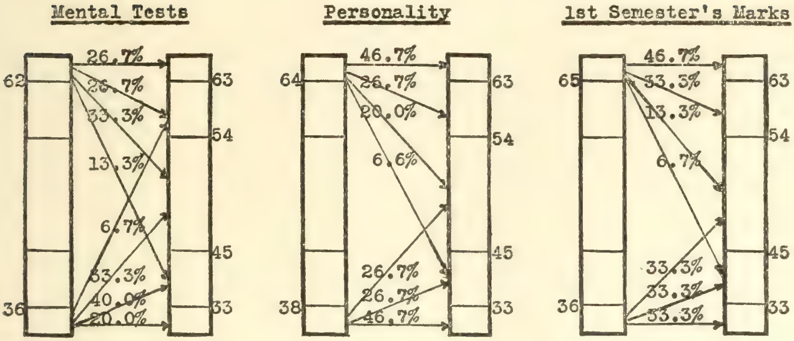


Fig. 15 Fig. 16 Fig. 17

PLATE 5. Levels of Achievement in Teaching-in-Training-School of Highest 10 Per Cent and Lowest 10 Per Cent.

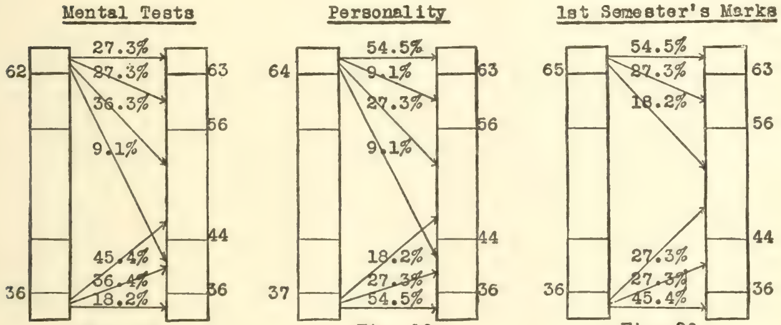


Fig. 18 Fig. 19 Fig. 20

PLATE 6. Levels of Achievement in Teaching-in-Field of Highest 10 Per Cent and Lowest 10 Per Cent.

The facts of Plates 3, 4, 5, and 6 are briefly summarized in the following table. The levels of achievement together with the per cents at each level may be readily seen from an inspection of the distributions for the highest and lowest tenths in Tests, Personality, and First-Semester's-Marks.

TABLE XXIII

LEVELS OF ACHIEVEMENT OF THE HIGHEST AND LOWEST TENTHS IN TESTS, PERSONALITY, AND FIRST SEMESTER MARKS

*See Below	Two Years' Marks	Teaching-in-Training-School	Normal-School-Success	In-Field-Success	%Group	"Grade" Group
Highest 10% in	40.0	26.7	46.7	27.3	Highest 10%	"A"
	33.3	26.7	20.0	27.3	Next 20%	"B"
Tests	26.7	33.3	33.3	36.3	Middle 40%	"C"
		13.3		9.1	Next 20%	"D"
	46.7	46.7	66.7	54.5	Highest 10%	"A"
Personality	33.3	26.7	13.3	9.1	Next 20%	"B"
	13.3	20.0	20.0	27.3	Middle 40%	"C"
	6.7	6.6		9.1	Next 20%	"D"
	60.0	46.7	60.0	54.5	Highest 10%	"A"
First-Semester-Marks	33.3	33.3	33.3	27.3	Next 20%	"B"
	6.7	13.3	6.7	18.2	Middle 40%	"C"
		6.7			Next 20%	"D"
Lowest 10% in						
	26.7	20.0	20.0	18.2	Lowest 10%	"E"
Tests	40.0	40.0	53.3	36.4	Next 20%	"D"
	33.3	33.3	26.7	45.4	Middle 40%	"C"
		6.7			Next 20%	"B"
	60.0	46.7	46.7	54.5	Lowest 10%	"E"
Personality	13.3	26.7	40.0	27.3	Next 20%	"D"
	26.7	26.7	13.3	18.2	Middle 40%	"C"
					Next 20%	"B"
	40.0	33.3	40.0	45.4	Lowest 10%	"E"
First-Semester-Marks	46.7	33.3	46.7	27.3	Next 20%	"D"
	13.3	33.3	13.3	27.3	Middle 40%	"C"
					Next 20%	"B"

*Of those "placed" among the highest ten per cent in Tests, 40 per cent achieved a place among the highest ten per cent ("A" group) in Two-Years'-Marks; 26.7 per cent in Teaching-in-Training-School; 46.7 per cent in Normal-School-Success; and 27.3 per cent in Teaching-in-Field, first year. In the next twenty per cent group were 33.3 per cent, 26.7 per cent, 20 per cent, and 27.3 per cent, respectively. And so on for the rest of the table.

(2) *As Determined by the Regression Equations for*

Two-Years'-Marks *on* Tests and Personality: (1)

$X_M = .251X_1 + .632X_2 + 5.85$

Teaching-in-Training School *on* Tests and Personality: (2)

$X_T = .233X_1 + .488X_2 + 13.95$

Normal-School-Success *on* Tests and Personality: (3)

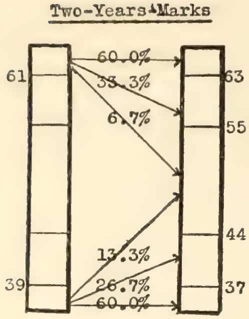
$X_N = .299X_1 + .602X_2 + 8.45$

In-Field-Success *on* Tests and Personality: (4)

$X_F = .146X_1 + .540X_2 + 15.7$

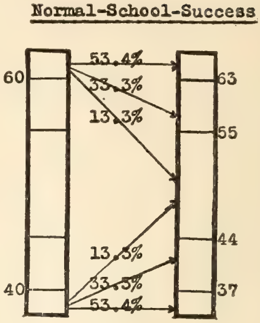
The accompanying graphs, Plate 7, indicate the levels of achievement, together with the per cents attaining to each level, in each of the four fields for the "highest" and "lowest" tenths in *Tests and Personality.* For example, Fig. 21, of the highest ten per cent (15 individuals) in *Tests and Personality* as determined by equation (1) above, 60% (or 9) of them were found in the *Two-Years'-Marks,* among the *highest* ten per cent (or "A" group); 33.3% (or 5) among the next twenty per cent (or "B" group); and 6.7% (or 1) among the middle forty per cent (or "C" group). None of them fell as low in school marks as the mean for the whole group. Of those in the lowest ten per cent (15 individuals) as determined by the same equation, 60% (or 9) of them were found to have place in the Two-Years'-Marks among the *lowest* ten per cent (or "E" group); 26.7% (or 4) among the next twenty per cent (or "D" group); and 13.3% (or 2) among the next forty per cent (or "C" group). But no one reached the mean for the whole group. No overlapping in achievement occurs here between individuals of the two groups, the highest and lowest tenths.

A similar interpretation applies to Figs. 22, 23, 24 for equations (2), (3), (4), above.



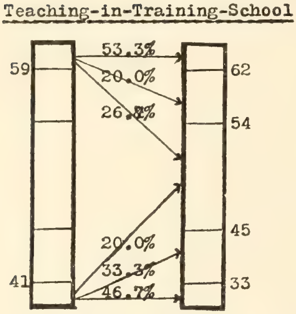
Two-Years'Marks

Fig. 21



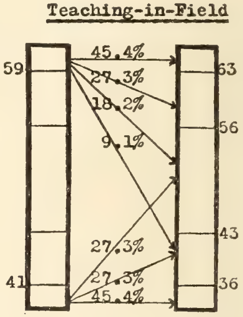
Normal-School-Success

Fig. 22



Teaching-in-Training-School

Fig. 23



Teaching-in-Field

Fig. 24

PLATE 7. Levels of Achievement of the Highest 10 Per Cent and the Lowest 10 Per Cent on the Basis of Tests and Personality Combined (Regression Values) in Two-Years'-Marks, Normal-School-Success, Teaching-in-Training-School, Teaching-in-Field.

(3) *As Determined by the Regression Equations for*

Two-Years'-Marks *on*
Tests, Personality and First-Semester-Marks: (1)

$$X_M = .106X_1 + .412X_2 + .420X_3 + 3.10$$

Teaching-in-Training-School *on*
Tests, Personality and First-Semester-Marks: (2)

$$X_T = .105X_1 + .306X_2 + .345X_3 + 12.2$$

Normal-School-Success *on*
Tests, Personality and First-Semester-Marks: (3)

$$X_N = .093X_1 + .406X_2 + .379X_3 + 6.10$$

In-Field-Success *on*
Tests, Personality and First-Semester-Marks: (4)

$$X_F = .033X_1 + .376X_2 + .311X_3 + 14.00$$

The accompanying graphs, Plate 8, indicate the levels of achievement together with the per cents attaining to each level, in each of the four fields for the *highest* and *lowest* tenths in Tests, Personality, and First-Semester-Marks. For example, Fig. 25, of the highest ten per cent (15 individuals) in the *combination* as determined by equation (1) above, 66.7% (or 10) of them were found in the Two-Years'-Marks among the highest ten per cent (or "A" group); 26.7% (or 4) among the next twenty per cent (or "B" group); and 6.6% (or 1) among the middle forty per cent (or "C" group. None of them fell as low as the mean for the whole group. Of those in the *lowest* ten per cent (15 individuals) as determined by the same equation, 73.3% (or 11) of them were placed in the Two-Years'-Marks among the lowest ten per cent (or "E" group); 20.0% (or 3) among the next twenty per cent (or "D" group); and 6.7% (or 1) among the middle forty per cent (or "C" group). But no one reached the mean for the group as a whole. No overlapping in achievement occurs here between members of the two sub-groups, highest and lowest tenths.

The same interpretation applies to Figs. 26, 27, 28 for equations (2), (3), (4), above.

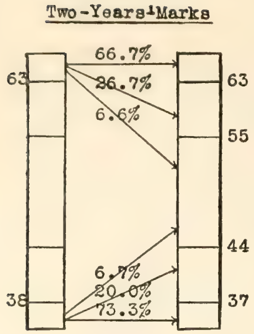


Fig. 25

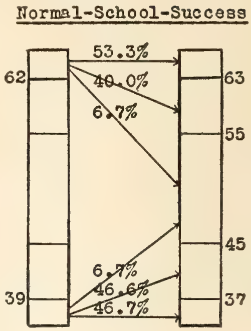


Fig. 26

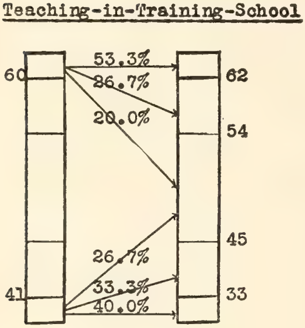


Fig. 27

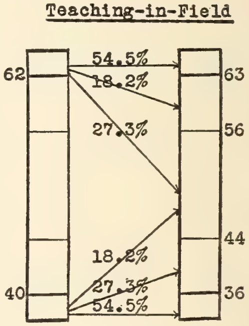


Fig. 28

PLATE 8. Levels of Achievement of Highest 10 Per Cent and Lowest 10 Per Cent on the Basis of Tests, Personality, and First-Semester-Marks Combined (Regression Values), in Two-Years'-Marks, Normal-School-Success, Teaching-in-Training-School, Teaching-in-Field.

The facts of Plates 7 and 8 are briefly summarized in the following table. The levels of achievement together with the per cents at each level may be readily seen from an examination of the distributions for the highest and lowest tenths in the combinations (regression values) of Tests and Personality, and Tests, Personality, and First-Semester-Marks.

TABLE XXIV

Levels of Achievement of the Highest and Lowest Tenths (as per Regression Values) for Tests and Personality; and Tests, Personality, and First Semester Marks

*See Below	Two-Years'-Grades	Teaching-in-Training-School	Normal-School-Success-	In-Field-Success	%Group	"Grade" Group
Highest 10% in	60.0	53.3	53.4	45.4	Highest 10%	"A"
Tests	33.3	20.0	33.3	27.3	Next 20%	"B"
and	6.7	26.7	13.3	18.2	Middle 40%	"C"
Personality				9.1	Next 20%	"D"
Tests						
Personality	66.7	53.3	53.3	54.5	Highest 10%	"A"
and	26.7	26.7	40.0	18.2	Next 20%	"B"
1st-Semester-	6.6	20.0	6.7	27.3	Middle 40%	"C"
Marks						
Lowest 10% in						
Tests	60.0	46.7	53.4	45.4	Lowest 10%	"E"
and	26.7	33.3	33.3	27.3	Next 20%	"D"
Personality	13.3	20.0	13.3	27.3	Middle 40%	"C"
Tests						
Personality	73.3	40.0	46.7	54.5	Lowest 10%	"E"
and	20.0	33.3	46.6	27.3	Next 20%	"D"
1st-Semester-	6.7	26.7	6.7	18.2	Middle 40%	"C"
Marks						

*Of those "placed" among the highest ten per cent in *Tests and Personality* (regression values), 60 per cent achieved position among the highest ten per cent (or "A" group) in Two-Years'-Grades; 53.3 per cent in Teaching-in-Training-School; 53.4 per cent in Normal-School-Success; and 45.4 per cent in Teaching-in-Field, 1st year. Of this same group, 33.3 per cent, 20 per cent, 33.3 per cent, and 27.3 per cent achieved position among the next twenty per cent (or "B" group) in these respectively; while 6.7 per cent, 26.7 per cent, 13.3 per cent, and 18.2 per cent won place among the middle forty per cent (or "C" group) respectively. In one case, Teaching-in-Field, 9.1 per cent dropped to the next twenty per cent (or "D" group).

Or you might read it in terms of chance, the per cents as so many chances per hundred. That is, for an individual whose performance entitles her to a place among the highest ten per cent in *Tests and Personality* the chances are about 60 to 100 that she will achieve position among the highest ten per cent (or "A" group) in Two-Years'-Grades, about 33 to 100 that she will be among the next twenty per cent (or "B" group), and about 7 to 100 that she will be among the middle forty per cent (or "C" group). There is not a chance of falling below this point in the two year's school marks.

With what degree of certainty, with what precision, can we forecast the achievement of the highest and lowest ten per cent of students both in the normal school and in the field their first year, basing our prediction on their standing (1) in Tests, Personality, and First-Semester-Marks separately? (2) in the combination of (*a*) Tests and Personality; (*b*) Tests, Personality, and First-Semester-Marks?

Given the problem of placing individuals in a group and told only that the range between the best and the worst is four units, and given a thousand guesses, the result would be on the basis of pure chance (or outright guessing) that a few of the highest 10 per cent (or lowest 10 per cent) would fall at each extreme (into both "A" and "E" groups); about twice this number into each twenty per cent group (or "B" and "D" groups); with a piling up of cases about the middle division (or "C" group). The average prediction error would be approximately 2.00.

On the basis of Tests, Personality, or First-Semester-Marks, singly or in combination (Tables XXIII and XXIV), the distribution is far different from what we should expect on the basis of pure chance. Instead of piling up about the middle, the larger numbers appear at or near the particular extreme which indicates correspondence between the measures and the different sorts of achievement. The larger the measure at its own extreme, the closer the correspondence.

TABLE XXV

PREDICTION ERROR* FOR THE EXTREME CASES—HIGHEST AND LOWEST TENTHS

1. ON THE BASIS OF TESTS, PERSONALITY, AND FIRST-SEMESTER-MARKS (SEPARATELY)

A Applies to the facts in Table XXIII	*Should read	Two-Years'-Grades	Teaching-in-Training-School	Normal-School-Success	In-Field-Success
	Highest 10% in				
	Tests	.87	1.33	.87	1.27
	Personality	.80	.89	.53	.81
	1st-Semester-Marks	.47	.80	.47	.64
	Lowest 10% in				
	Tests	1.07	1.27	1.07	1.27
	Personality	.67	.80	.67	.64
	1st-Semester-Marks	.73	1.00	.73	.82

2. ON THE BASIS OF (*a*) TESTS AND PERSONALITY; (*b*) TESTS, PERSONALITY, AND FIRST-SEMESTER-MARKS

B Applies to the facts in Table XXIV	*Should read	Two-Years'-Grades	Teaching-in-Training-School	Normal-School-Success	In-Field-Success
	Highest 10% in				
	Tests and Personality	.47	.73	.70	.91
	Tests, Personality, and 1st-Semester-Marks	.40	.67	.53	.73
	Lowest 10% in				
	Tests and Personality	.53	.73	.60	.82
	Tests, Personality, and 1st-Semester-Marks	.33	.87	.50	.64

*Prediction error is obtained by multiplying the percentage displacement in each case by the standard deviation (or a multiple) and summing. From the mid-point of "A" (Table XXIII) to the mid-point of "B" is approximately 1 S.D.; from the mid-point of "A" to the mid-point of "C" is approximately 2 S.D.; and so on; so that .87 in upper left compartment is obtained by summing

$$(33.3\% \times 1) + (26.7\% \times 2). = .87$$

No less striking is the reduction in amount of error as shown in Table XXV above. Tests alone would reduce the error more than 40 per cent; Personality alone more than 60 per cent; and First-Semester-Marks alone more than 60 per cent. Tests and Personality together would reduce it about 65 per cent, and the three together would reduce it some 70 per cent or more; the former giving an average misplacement of about 35 per cent of that given by pure chance, the latter a little less than 30 per cent.

With what differing degrees of certainty can we forecast this achievement on the basis of (1) data available at the beginning of the training period or entrance into the normal school, and (2) data which become available at the end of the first semester's work? That is, how much more valuable are first semester grades than test scores and personality estimates for this purpose, and how much value do they add to the combination?

An inspection of Table XXIII shows the percentage distribution of the highest and lowest tenths in Tests, Personality, and First-Semester-Marks among the different levels of achievement in (1) Two-Years'-Marks, (2) Teaching-in-Training-School, (3) Normal-School-Success, and (4) Teaching-in-Field, first year. These distributions indicate but little difference between Personality and First-Semester-Marks, both being superior to Tests. Comparing these two as to the "placement" of the highest 10 per cent among the different groups, there is practically no difference in the "A" groups (in two cases, the same percentages; in one of the other two cases, Personality is higher with 66.7 to 60.0; in the other, it is lower with 46.7 to 60.0). In the "B" groups, Personality shows up better in all except one case, Two-Years'-Marks in which it is the same. It is to be noted that the presence of high numbers outside of the first row indicates inferiority in value. In the "C" group, it is inferior in all four cases. In the "D" group, it is inferior in two, and the same in the other two. Similarly the distributions are shown for the lowest 10 per cent.

Table XXVA further affirms these facts; the smaller the number, the less the error, and the better the prediction. The error of prediction is a little larger in connection with the highest 10 per cent for Personality but smaller in all types of success except one, Two-Years'-Marks, in which it is the same. It would seem, therefore, that little of worth has been added by waiting for the first half year's grades alone and that practically as close a prediction is possible on the basis of Personality.

A like interpretation of Table XXIV shows that, when First-Semester's-Marks are combined with Tests and Personality, the prediction is somewhat closer than that given by a combination of those two alone. The two in combination, however, yield a prediction appreciably closer than First-Semester-Marks alone. The numbers for the "A" group are larger generally, and correspondingly smaller for the "B" and "C" groups.

Table XXVB would seem to indicate that the difference is not a large one, although perceptible in all types of success with the exception of Teaching-in-Training-School (lowest 10 per cent). In that the error is somewhat *larger!* On the whole, however, the error is reduced when the first half year's grades are added—the reduction being on the average approximately one-tenth (.10). Although these marks apparently give an added value for predicting achievement in the field for the extreme variants, the added increment as found for the group as a whole (110 individuals) is not large, and for practical purposes guidance could be based on the facts of Tests and Personality in the large majority of cases.

THE ORIGIN OF THE EXTREME VARIANTS

(1) *As Shown by Level of Achievement Compared with Position in Measures.* Plates 9 to 12, indicate the sources of supply —both levels of position and per cents from each level—of those individuals composing the *highest* and *lowest* tenths in each of the four fields of achievement,—Two-Years'-Marks, Teaching-in-Training-School, Normal-School-Success, and Teaching-in-Field, first year. For example, of the *highest* ten per cent (15 individuals) in Two-Years'-Marks, 93.3% (or 14) were drawn from the upper half of the distribution in Tests: 40% (or 6) Fig. 29 below from the highest ten per cent (or "A" group; 40% (or 6) of them were among the next twenty per cent (or "B" group); and 20% (or 3) were among the middle forty per cent (or "C" group). Only one was below the mean of the whole group.

Of the *lowest* ten per cent (15 individuals), 86.7% (or 13) came from the lower half in Tests: 20% (or 3) of them were among the lowest ten per cent (or "E" group); 46.7% (or 7) among the next twenty per cent (or "D" group); and 33.3% (or 5) among the middle forty per cent (or "C" group). Only two individuals had a position in the Tests above the average of the whole group, and they but slightly above.

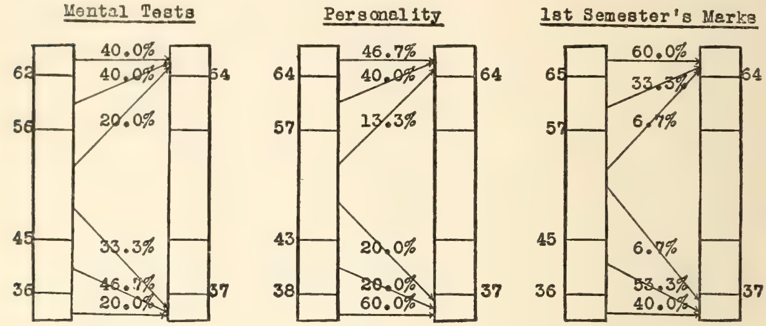


Fig. 29 Fig. 30 Fig. 31

PLATE 9. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Two-Years'-Marks.

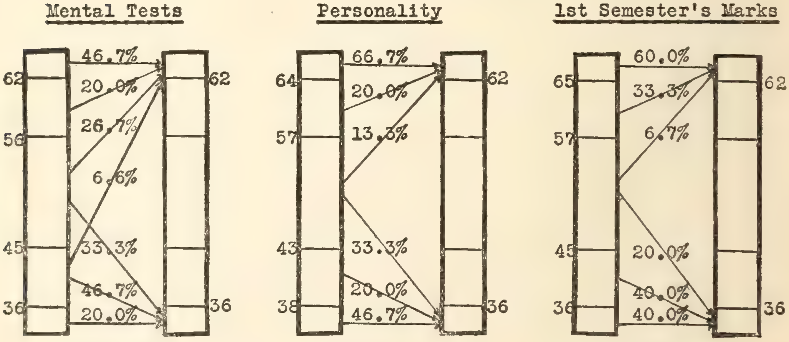


Fig. 32 Fig. 33 Fig. 34

PLATE 10. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Normal-School-Success.

Similar interpretations apply to graphs indicating positions of the individuals composing the same groups when drawn from the distributions for Personality and First-Semester-Marks; also, Plates 10, 11, and 12 showing the sources supplying the highest and lowest tenths in Teaching-in-Training-School, Normal-School-Success, and Teaching-in-Field, first year.

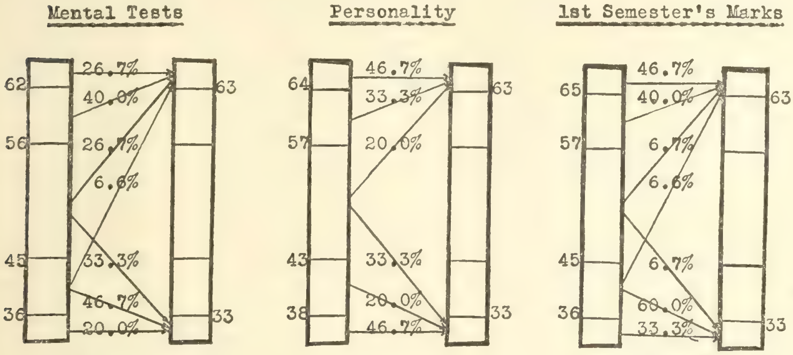


Fig. 35 Fig. 36 Fig. 37

PLATE 11. Levels Which Supply Highest 10 Per Cent and Lowest 10 Per Cent in Teaching-in-Training-School.

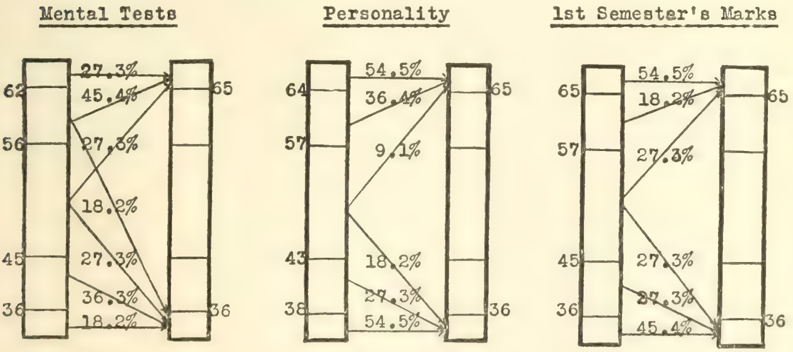


Fig. 38 Fig. 39 Fig. 40

PLATE 12. Levels Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent in Teaching-in-Field.

TABLE XXVI

PER CENTS OF THE HIGHEST AND LOWEST TENTHS IN ACHIEVEMENT AS DRAWN FROM THE DIFFERENT LEVELS IN TESTS, PERSONALITY, FIRST-SEMESTER-MARKS

*See below	Tests	Personality	First Semester Marks	% Group	"Grade" Group
Highest 10% in Two-Years'-Grades	40.0	46.7	60.0	Highest 10%	"A"
	40.0	40.0	33.3	Next 20%	"B"
	20.0	13.3	6.7	Middle 40%	"C"
				Next 20%	"D"
Highest 10% in Teaching-in-Training-School	26.7	46.7	46.7	Highest 10%	"A"
	40.0	33.3	40.0	Next 20%	"B"
	26.7	20.0	6.7	Middle 40%	"C"
	6.6		6.6	Next 20%	"D"
Highest 10% in Normal-School-Success	46.7	66.7	60.0	Highest 10%	"A"
	20.0	20.0	33.3	Next 20%	"B"
	26.7	13.3	6.7	Middle 40%	"C"
	6.6			Next 20%	"D"
Highest 10% in In-Field-Success First Year	27.3	54.5	54.5	Highest 10%	"A"
	45.4	36.4	18.2	Next 20%	"B"
	27.3	9.1	27.3	Middle 40%	"C"
				Next 20%	"D"
Lowest 10% in Two-Years'-Grades	20.0	60.0	40.0	Lowest 10%	"E"
	46.7	20.0	53.3	Next 20%	"D"
	33.3	20.0	6.7	Middle 40%	"C"
				Next 20%	"B"
Lowest 10% in Teaching-in-Training-School	20.0	40.0	33.3	Lowest 10%	"E"
	46.7	26.7	60.0	Next 20%	"D"
	33.3	33.3	6.7	Middle 40%	"C"
				Next 20%	"B"
Lowest 10% in Normal-School-Success	20.0	46.7	40.0	Lowest 10%	"E"
	46.7	20.0	40.0	Next 20%	"D"
	33.3	33.3	20.0	Middle 40%	"C"
				Next 20%	"B"
Lowest 10% in In-Field-Success First Year	18.2	54.5	45.4	Lowest 10%	"E"
	36.3	27.3	27.3	Next 20%	"D"
	27.3	18.2	27.3	Middle 40%	"C"
	18.2			Next 20%	"B"

*Of those whose achievement places them among the highest ten per cent in Two-Years'-Marks, 40% of them were ranked in the highest ten per cent (or "A" group) in Tests, 40% in the next twenty per cent (or "B" group), and 20% in the middle forty per cent (or "C" group). In Personality, 46.7% of them had place in the highest ten per cent (or "A" group), 40% in the next twenty per cent (or "B" group), and 13.3% in the middle forty per cent (or "C" group). In First-Semester-Marks, 60% of them were in the highest ten per cent (or "A" group), and 33.3% in the next twenty per cent (or "B" group), and 6.7% in the middle 40% (or "C" group).

Similar reading of the rest of the table.

(2) *As Determined by the Regression Equations for—*

Two-Years'-Marks *on* (1)

(*a*) Tests and Personality:

$$X_M = .25X_1 + .63X_2 + 6.0$$

(*b*) Tests, Personality and First-Semester-Marks:

$$X_M = .106X_1 + .412X_2 + .420X_3 + 3.1$$

Teaching-in-Training-School *on* (2)

(*a*) Tests and Personality:

$$X_T = .23X_1 + .49X_2 + 14.0$$

(*b*) Tests, Personality and First-Semester-Marks:

$$X_T = .105X_1 + .306X_2 + .345X_3 + 12.2$$

Normal-School-Success *on* (3)

(*a*) Tests and Personality:

$$X_N = .229X_1 + .602X_2 + 8.45$$

(*b*) Tests, Personality and First-Semester-Marks:

$$X_N = .093X_1 + .406X_2 + .379X_3 + 6.1$$

Teaching-in-Field, First Year, *on* (4)

(*a*) Tests and Personality:

$$X_F = .146X_1 + .540X_2 + 15.7$$

(*b*) Tests, Personality and First-Semester-Marks:

$$X_F = .033X_1 + .376X_2 + .311X_3 + 14.0$$

Plates 13 and 14 indicate the levels of the sources (*Tests and Personality,* and *Tests, Personality, and First-Semester-Marks,*) composing the *highest* and *lowest* tenths in each of the four fields (Two-Years'-Marks, Teaching-in-Training-School, Normal-School-Success, and In-Field-Success, First Year) together with the per cents coming from each level. For example, the highest ten per cent (15 individuals) in Two-Years'-Grades were found to have been supplied by the upper half of the distributions as determined by equations (1*a*) and (1*b*) above. In the first case, Fig. 41, 60% (or 9) of them came from the *highest* ten per cent (or "A" gorup) in Tests and Personality; 33.3% (or 5) from the next twenty per cent (or "B" group); and 6.7% (or 1) from the middle forty per cent (or "C" group); but none from below the mean of the whole group. In the second case, Fig. 45, Plate 14, 66.7% (or 10) of them came from the *highest* ten per cent (or "A" group) in Tests, Personality, and First-Semester-Marks; and 20.0%

(or 3) from the next twenty per cent (or "B" group); 13.3% from the middle forty per cent (or "C" group); none from below the mean of the whole group.

The *lowest* ten per cent (15 individuals) in Two-Years'-Marks were found likewise to have been supplied by the lower half of the distributions as determined by these same equations. In the first case, 60% (or 9) of the individuals came from the *lowest* ten per cent (or "E" group) in Tests and Personality; 40% (or 6) from the next twenty per cent (or "D" group); none from above the

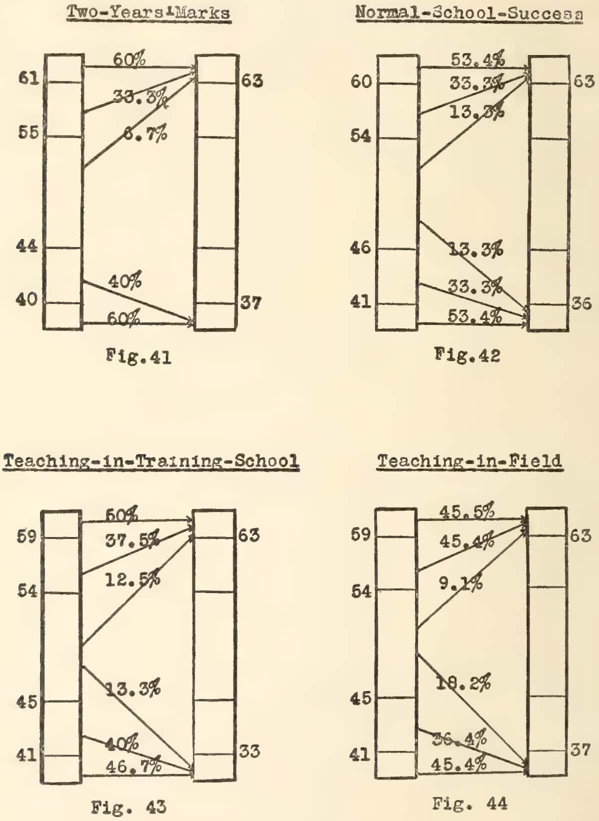


Fig.41 Fig.42 Fig. 43 Fig. 44

Plate 13. Levels on the Basis of Tests and Personality Combined (Regression Values), Which Supply the Highest 10 Per Cent and Lowest 10 Per Cent for Two-Years'-Marks, Normal-School-Success, Teaching-in-Training School, Teaching-in-Field.

thirty-per-cent-from-the-lowest point. In the second case, Fig. 45, 73.3% (or 11) of them came from the lowest ten per cent (or "E" group) in Tests, Personality, and First-Semester-Marks; 20% (or 3) from the next twenty per cent (or "D" group); and 6.7% (or 1) from the middle forty per cent (or "C" group); one case at the mean for the whole group; none above the mean, and no overlapping between the two groups, either in standing in *Tests and Personality* or in *Tests, Personality, and First-Semester-Marks,* composing the highest and lowest tenths in the Two-Years'-Marks.

Similar interpretations of Figs. 42, 43, 44 and 46, 47, 48 for equations (2*a*) and (2*b*), (3*a*) *and* (3*b*), and (4*a*) and (4*b*).

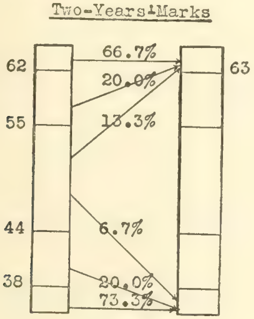


Fig. 45

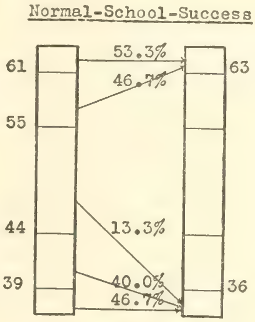


Fig. 46

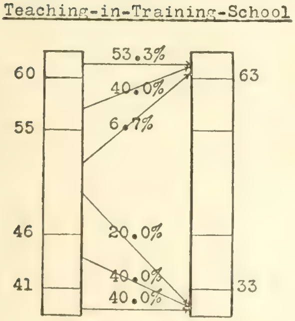


Fig. 47

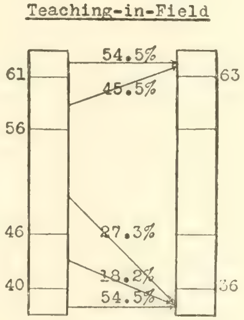


Fig. 48

Plate 14. Levels on the Basis of Tests, Personality, and First-Semester-Marks Combined (Regression Values), Which Supply the Highest 10 Per Cent and Lowest 10 Per cent for Two-Years'-Marks, Normal-School-Success, Teaching-in-Training-School, Teaching-in-Field.

TABLE XXVII

PER CENTS OF THE HIGHEST AND LOWEST TENTHS IN ACHIEVEMENT AS DRAWN FROM THE DIFFERENT LEVELS IN TESTS AND PERSONALITY, AND TESTS, PERSONALITY, AND FIRST-SEMESTER-MARKS AS PER THEIR REGRESSION VALUES

*Should read	Tests and Personality	Tests, Personality, and First-Semester-Marks	% Group	"Grade" Group
Highest 10% in Two-Years'-Marks	60.0	66.7	Highest 10%	"A"
	33.3	20.0	Next 20%	"B"
	6.7	13.3	Middle 40%	"C"
			Next 20%	"D"
Teaching-in-Training-School	50.0	53.3	Highest 10%	"A"
	37.5	40.0	Next 20%	"B"
	12.5	6.7	Middle 40%	"C"
			Next 20%	"D"
Normal-School-Success	53.4	53.3	Highest 10%	"A"
	33.3	46.7	Next 20%	"B"
	13.3		Middle 40%	"C"
			Next 20%	"D"
In-Field-Success, First Year	45.5	54.5	Highest 10%	"A"
	45.4	45.5	Next 20%	"B"
	9.1		Middle 40%	"C"
			Next 20%	"D"
Lowest 10% in Two -Years'-Marks	60.0	73.3	Lowest 10%	"E"
	40.0	20.0	Next 20%	"D"
		6.7	Middle 40%	"C"
			Next 20%	"B"
Teaching-in-Training-School	46.7	40.0	Lowest 10%	"E"
	40.0	40.0	Next 20%	"D"
	13.3	20.0	Middle 40%	"C"
			Next 20%	"B"
Normal-School-Success	53.4	46.7	Lowest 10%	"E"
	33.3	40.0	Next 20%	"D"
	13.3	13.3	Middle 40%	"C"
			Next 20%	"B"
In-Field-Success, First Year	45.4	54.5	Lowest 10%	"E"
	36.4	18.2	Next 20%	"D"
	18.2	27.3	Middle 40%	"C"
			Next 20%	"B"

*Of those who achieve positions among the highest ten per cent in Two-Years'-Marks (Teaching-in-Training-School, etc.), 60% were ranked among the highest ten per cent (or "A" group) in Tests and Personality, 33.3% among the next twenty per cent (or "B" group), and 6.7% among the middle forty per cent (or "C" group). In Tests, Personality, and First-Semester-Marks, 66.7% were ranked among the highest ten per cent (or "A" group), 33.3% among the next twenty per

TABLE XXVIII

Prediction Error for the Extreme Cases—Highest and Lowest Tenths—In Achievement as to the Sources from Which They Came

1. Tests, Personality, and First-Semester-Marks (Separately)

A

*See below	Highest 10%			Lowest 10%		
	Tests	Personality	First-Semester-Marks	Tests	Personality	First-Semester-Marks
Two-Years'-Marks	.80	.63	.47	1.13	.60	.67
Teaching-in-Training-School	1.07	.73	.73	1.13	.87	.73
Normal-School-Success	.93	.37	.48	1.23	.87	.80
In-Field-Success	1.00	.55	.73	1.45	.64	.82

2. Combination of (*a*) Tests and Personality; (*b*) Tests, Personality, and First-Semester-Marks

B

*See below	Highest 10%		Lowest 10%	
	Tests and Personality	Tests, Personality, and 1st-Semester-Marks	Tests and Personality	Tests, Personality and 1st-Semester-Marks
Two-Years'-Marks	.47	.47	.40	.33
Teaching-in-Training-School	.63	.53	.67	.80
Normal-School-Success	.60	.47	.60	.67
In-Field-Success	.64	.46	.73	.73

*Prediction error is obtained in the same way as explained on page 103. As used here, it means that, should we guess the highest (or lowest) ten per cent in each field of achievement came from the highest (or lowest) ten per cent in Tests, Personality, etc., the error in each case would be as indicated in the corresponding compartment. That is, by "placing" all of the highest 10% in Two-Years'-Marks back among the highest 10% in Tests, we make an average error in placement of .80 (of one S.D.), .8 of one grade ("A," "B," "C," etc.)

---

cent (or "B" group), 13.3% among the middle 40% (or "C" group), and none below the mean of the whole group. Similar interpretation for the rest of the table.

Or you might read it in terms of chance, the per cents as so many chances per hundred. That is, for one whose achievement places her among the highest ten per cent in Two-Years'-Marks, the chances are about 67 to 100 that she ranked among the highest ten per cent (or "A" group) in *Tests and Personality;* about 27 to 100 that she ranked among the next twenty per cent (or "B" group); and about 7 to 100 that she ranked among the middle forty per cent (or "C" group). Likewise, the chances are about 67 to 100 that she ranked among the highest ten per cent (or "A" group) in *Tests, Personality, and First-Semester-Marks;* about 20 to 100 that she ranked among the next twenty per cent (or "B" group); with about 13 chances in 100 of being in the middle forty per cent (or "C" group).

To what extent does high achievement in school grades and in teaching indicate high standing in Tests and Personality, and to what extent does the reverse hold true,—poor achievement in grades and in teaching imply low standing in Tests and Personality? To what extent would we be justified in assuming, for those who rank high in achievement, correspondingly high position in Tests and Personality,—and the reverse for those standing low in achievement?

An examination of Figs. 29 to 48, Plates 9 to 14, and an inspection of Tables XXVI, XXVII, and XXVIII, show a very different relationship from what would be expected by pure chance. Large numbers with extreme grades (highest 10 per cent and lowest 10 per cent, "A" 's and "E" 's) indicate close correspondence—the larger the percentages in the extreme ranges and the smaller in the middle ranges the closer the correspondence; and the smaller the numbers in Table XXVIII, the smaller the error, and consequently better the prediction.

The facts here, Table XXVI and Table XXVIIIA, indicate clearly that we should make a smaller error to assume for high or low achievement a like position in Personality and First-Semester-Marks than in Tests, although an appreciable correspondence is apparent even there. Relatively large numbers are seen for the highest 10 per cent and lowest 10 per cent among the "A" group and "E" group respectively, and numbers around 1.00 in Table XXVIIIA show the chance error reduced by half; for Personality and First-Semester-Marks the error is reduced by about two-thirds.

To what extent may we make a similar assumption to the above, for a combination of *Tests and Personality,* and *Tests, Personality, and First-Semester-Marks?* The assumption of correspondence between achievement and either of these combinations would be justified in larger measure than in the case of any one of them alone, and somewhat closer to the facts for the combination of the three than for only two of them. Just how much closer either of these combinations would prove may be seen by comparing each of the distributions (column-wise) in Table XXVII with the three sets of similar distributions in Table XXVI,—particularly the first per cents in each compartment with the numbers in the corresponding compartments, the second per cents in each compartment with the numbers in the corresponding compartments, and so on. Similarly, compare *Part B* with *Part A* (Table XXVIII), and the magnitude of the error will, in practically every case, be shown to be smaller in the former than in the latter.

Furthermore, a comparison of columns 1 and 2 in Table XXVII and the two parts of Table XXVIIIB will show that the facts in the combination of the three (*Tests, Personality, and First-Semester-Marks*) are somewhat closer to the assumption than are the facts of the combination of only two (*Tests and Personality*).

JUSTICE AND INJUSTICE IN ELIMINATION

(1) *Of the "Final" Low Grades.* To what extent would injustice be done to deserving individuals by advising all those falling below a certain position (a critical score) in Tests, Personality, and First-Semester-Marks, and "combinations" of them, to withdraw from the normal school and to apply their "talents" to some other type of activity? And to what extent would justice be done those destined to achieve positions below the "satisfactory limit" in school and in the field?

Assuming *complete failure* for those whose "achievement" places them as far as 2 S.D. below the mean (average for the group) and *practical failure* for all beyond 1.5 S.D. below the mean (and there are good reasons for this assumption both theoretically as to the nature of the normal distribution placing only some 2.27 per cent and 6.68 per cent respectively below these points, and practically as a matter of fact on the basis of the data in this study placing 3 per cent and 8 per cent in school success and 2 per cent and 8 per cent in success in field the first year below these same points respectively), in how far should we be justified in making eliminations on the basis of data available at the beginning of the normal school course and at the end of the first semester's work?

There are then (*a*) 4 *complete* and 15 *practical* failures in the normal school; (*b*) 2 *complete* and 9 *practical* failures in the field, first year. Tables XXIX and XXX below indicate the results which would be obtained in each case of elimination on the different bases,—the numbers (or per cents) advised out, together with the justice or injustice done.

An inspection of Table XXIXA and B shows how many of the *complete* and *practical* failures would be eliminated on the basis of Tests, Personality, and First-Semester-Marks, and combinations of them; also the number likewise affected whose records in school and in the field place them somewhere between the borderland of failure and average success.

TABLE XXIX

ELIMINATION OF FAILURES IN NORMAL SCHOOL WORK

(N = 4 (— 2 S.D.)
(N = 15 (— 1.5 S.D.)

A

When Eliminated on the Basis of	Numbers Eliminated		Number of Cases in Which the Result Is									
			Complete Justice (1)		Reasonable Justice (2)		Doubtful Justice or Injustice (3)		Slight Injustice (4)		Appreciable Injustice (5)	
	—2 S. D.	—1.5 S. D.										
Tests	5	12	0	1	1	1	0	2	3	5	1	3
Personality	4	11	1	4	2	3	1	2	0	0	0	2
First-Semester-Marks	6	14	2	3	3	3	0	3	1	2	0	3
Normal-School-Success *on* Tests and Personality	5	15	2	4	1	5	1	3	0	1	1	2
Normal-School-Success *on* Tests, Personality and First-Semester-Marks	6	16	2	4	3	4	1	4	0	3	0	1
Weighted—Tests and Personality	0	12	0	3	0	4	0	2	0	2	0	1

Numbers in 2nd half of each column include numbers in 1st half.

ELIMINATION OF FAILURES IN FIELD WORK (1ST YEAR)

(N = 2 (—2 S.D.)

B (N = 9 (—1.5 S.D.)

When Eliminated on the Basis of	Numbers Eliminated		Number of Cases in Which the Result Is									
			Complete Justice (1)		Reasonable Justice (2)		Doubtful Justice or Injustice (3)		Slight Injustice (4)		Appreciable Injustice (5)	
	—2 S. D	—1.5 S. D										
Tests	4	9	0	0	0	1	0	2	2	4	2	2
Personality	4	9	1	2	1	4	0	1	0	0	1	2
First-Semester Marks	5	11	2	2	2	3	0	1	1	4	0	1
In-Field-Success on Tests and Personality	2	10	1	2	1	3	0	1	0	2	0	2
In-Field-Success on Tests, Personality, and 1st-Semester-Marks	3	11	1	2	1	4	0	2	1	2	0	1
Weighted—Tests and Personality	0	11	0	2	0	3	0	2	0	2	0	2

Under the column headed "Complete Justice," the numbers represent in each case the *complete* failures who would thus be advised out on the basis indicated. The first half of the compartment shows the number so affected when we used only those cases who fall below −2 S.D., the second half the number so affected when we use all cases falling below −1.5 S.D. Add to this the column headed "Reasonable Justice," and you have all failures, complete and practical, who would be eliminated in the same manner as just stated. Column 3 may be passed up as containing cases of borderline indetermination. Columns 4 and 5 show the numbers which would be thrown out on the basis of each set of measures, whose achievement in school or field places them between doubtful and low-average-success, in the upper "D" and lower "C" groups in achievement.

Using *Tests* alone as our basis, we find five cases below –2 S.D., none of whom were complete failures in school and only one practical failure. To rule out of school on this basis would result in slight injustice to three individuals and unquestioned injustice to one another. There are twelve cases below –1.5 S.D. (including the five cases below –2 S.D.), one of whom ranks as a complete failure in school work, another as a practical failure, two of doubtful status, and eight (5 plus 3) as more or less successful. To rule out on this basis would result in justice to two individuals and injustice to eight others.

Of the four complete failures in normal school work, none would be eliminated by taking –2 S.D. as the critical point in Tests, and only one case if –1.5 S.D. be chosen. Of the fifteen practical failures (including the four complete failures) in school work, only one individual would be dropped by reason of being below –2 S.D. in Tests, and one other below –1.5 S.D. Then but one of the four complete failures could be reached by this means and only two of the twelve practical failures. Our justice ratio would be 2/10 (or 20%); efficiency ratio, 2/15 or 13%). There seems to be no absolutely safe *critical score* here.

Using *Personality* alone we find four cases below –2 S.D., one of whom achieved complete failure in school, two others practical failure, and one of doubtful status; none above. We find twelve cases (including these four) below –1.5 S.D., four of whom were complete failures in school, three others practical failures, two of doubtful status, and two successful. Justice would result in seven cases, injustice in two.

Of the four complete failures in school, one would be eliminated by taking –2 S.D. as the critical point in Personality; all four, if –1.5 S.D. be chosen; and of the fifteen practical failures (including the four), three would be eliminated by reason of being below –2 S.D. in Personality, and four others 4 minus 1 and 3 minus 2 being below –1.5 S.D. All of the complete failures could thus be reached, and seven of the 15 practical failures. Justice ratio is 7/9 (or 78%); efficiency ratio, 7/15 (or 47%).

Determined in like manner for First-Semester-Marks, these ratios are 60% (6/10) and 40% (6/15), respectively.

For Tests and Personality (regression values), they are 75% (9/12) and 60% (9/15).

For Tests, Personality, and First-Semester-Marks (regression values), they are 67% (8/12) and 53% (8/15).

For Tests and Personality (weighted 1 and 3, respectively) they are 70% (7/10) and 47% (7/15).

From a like treatment of the facts of Table XXIXB, we obtain justice and efficiency ratios for the different measures as follows:

(1) Tests....14% (1/7) and 11% (1/9).
(2) Personality....67% (6/9) and 67% (6/9).
(3) First-Semester-Marks.....50% (5/10) and 56% (5/9).
(4) Tests and Personality.....56% (5/9) and 56% (5/9).
(5) Tests, Personality, and First-Semester-Marks....67% (6/9) and 67% (6/9).
(6) Tests and Personality (weighted)......56% (5/9) and 56% (5/9).

Note: The critical point for these was taken at –1.5 S.D.

The magnitude of the paired ratios denotes the value of the set of measures as a basis for elimination. *Tests and Personality* in Table XXIX A (for school work), and *Tests, Personality, and First-Semester-Marks* in XXIX B (for field work) make the best showing. Neither pair of ratios is sufficiently large, however, to justify rigid application for practical purposes. Too many cases of injustice occur for the number favorably affected. It is apparent then that –1.5 S.D. is too far up the scale, is above the critical zone. If –2 S.D. is taken as the critical point, the justice ratio is increased for practically all measures and in some cases very materially increased,—particularly *Tests and Personality* and *Tests, Personality, and First-Semester-Marks* in B, increasing to 100 per cent in each case. But at the same time the efficiency ratios are generally reduced, and the numbers affected are greatly reduced.

(2) *Of Those Who Dropped Out.* Of the eighteen individuals who dropped out of school before having completed as much as a year's work "causes" were ascertained for thirteen cases. Poor work and consequent discouragement were the prime moving forces with one group (Nos. 1, 2, 3); outside interests, ranging from momentary impulses and fancy to marriage, in six cases (Nos. 4, 5, 6, 7, 8, 9); poor health and illness accounted for three others (Nos. 10, 11, 12); and financial difficulties given by another (No. 13). No definite reasons were given and the causes could not be

determined for the other five cases all of whom did low-average-work when compared with the performance of the entire class. Table XXX below shows to what extent this group could have been "reached" on the basis of Test-scores and Personality-estimates singly and in combination.

TABLE XXX

Elimination of Those Who Withdrew

When eliminated on the basis of	Numbers Reached and Results			
	—2 S. D. Critical Point		—1.5 S.D. Critical Point	
	Justice	Injustice	Justice	Injustice
Tests	3	1	4	2
Personality	2	0	3	0
Regression Values for				
(1) Normal School Success on Tests and Personality	5	1	6	3
(2) In-Field-Success on Tests and Personality	2	0	6	0

Note: "Justice" and "Injustice" here have reference to the pairs of factors, *poor-work* and *outside-interests,* over which the individual has control and for which she is largely responsible on the one hand, and *ill-health* and *financial-stress,* over which she does not have such control and for which she is not responsible on the other.

If we consider the thirteen individuals as two somewhat different and distinct groups from the standpoint of "causes" for withdrawing—those whose abilities or inclinations, or both, prevented them from meeting the institutional requirements and those whose physical condition or finances *may* have been the chief cause of the failure—nine in the former group and four in the latter—practically every one of them could have been eliminated at the very outset on the basis of test-ability and personality, by making use of the values obtained from the regression equation for predicting achievement in the normal school work. This equation is $X_N = .229X_1 + .602X_2 + 8.45$.

If –2 S.D. be taken as the *critical point,* five of the first group and one of the second could have been advised out. If –1.5 S.D.

be chosen, six of the first and three of the second could have been eliminated. No serious (real) injustice would have resulted, and health might have been conserved in three cases. The efficiency ratio being 69% (9/13) with a justice ratio 100% (9/9), if health be a value worthy of our best efforts at safeguarding and preserving. Otherwise, this ratio drops to 67% (6/9).

By combining these facts with the corresponding facts in Table XXIXA the following results would have been possible at the very outset:

1. With –2 S.D. as the critical point, eleven individuals would have been eliminated; eight of whom either dropped out of school because of poor work (lack of progress) or failed finally to achieve position above the lowest 10% of the group; one of whom had place among the next 6% in school achievement; another won final place in school work slightly below the average of the whole group; and one withdrew before the end of the first year because of ill health. In only one case (out of eleven) would appreciable injury have been done, and her performance in the training course registered but average (or "C") grade. Considered from the standpoint of accomplishment or the lack of it, of the thirteen cases of failure (the nine who dropped out because of "outside-interests" and "poor-work," and the four whose achievement placed them below –2 S.D. in Normal-School-Success) seven could have been disposed of at the outset.
2. With –1.5 S.D. as the critical point twenty-four individuals would have been eliminated; fifteen of whom either dropped out on account of poor work (lack of progress) before the end of the first year or failed finally to achieve position above the lowest 10% of the group; three of whom had place among the next 6% in school achievement; three slightly below the average, and three others dropped out during the first year because of illness. In only three cases would injustice have been done—mild in one case, appreciable in two. And of the twenty-four failures (the nine who withdrew the first year because of unsatisfactory work and the fifteen whose final achievement in normal-school-work placed them below –1.5 S.D.) fifteen could have been given proper guidance at the outset.

A safe and sane policy, it would seem, to insure maximum justice and efficiency, might follow from the fore-going facts and figures.

A critical-zone should be set at –2 S.D., and a near-critical-zone between this point and –1.5 S.D; the former to be used as a basis for possible elimination; the latter, for further and fuller investigation of the individual case and for purposes of acquainting the subject with her possibilities,—the levels of achievement possible and the probabilities of attaining to those levels. The data do not justify stronger suggestions or recommendations.

CHAPTER VII

CONCLUSIONS

THE METHOD

Questionnaire methods have increasingly small place in problems of scientific guidance, educational and vocational. False memories and deep-rooted present purposes are too often misinterpreted for past bents and related abilities. Likes and dislikes, aims and ambitions, attitudes and aptitudes, efforts and energies, of yesterday cannot be adequately measured by the achievements and accumulations of to-day or to-morrow, however indicative and significant they may be. Factors impossible of precise evaluation influence the outcome and render every "present" situation an inexact, and frequently very inaccurate, index of past proclivities and potentialities. To determine definitely the contents of a reservoir at a particular date, measurements and analyses must be made in season, —before the sun's rays, climatic shifts, organic and inorganic changes, and other unnumbered influences have registered their effects. Likewise, to discover the feelings and thoughts of an individual, diagnosis must be made of the experience itself and not of its "impression,"—memory and fact are frequently synonymous, and as frequently not.

This thought has been uppermost in mind throughout the investigation. To determine relationship between certain types of achievement and "personal" qualities which will serve the purposes of prognosis and practical guidance, measurements of the same individuals at different stages are necessary,— at the time when guidance and direction are needed, and when achievement is effected. Only by comparing facts and measurements of the one with those found to obtain for the other, can actual relationship be determined. To make many different sets of measurements and at many different times would be superior to making two,—and the larger the number the better the results.

Most studies so far reported have been approached in the reverse order,—present conditions and reactions and abilities compared with past "records" and different footrules employed for measuring similar performance at the two stages. Because of this difference in purpose, method, and means employed, the present study is not comparable to other related investigations in the same general field. Knight's[1] study was concerned with the measurement of certain qualities present in the teacher-in-service,—post-diagnostic rather than prognostic. Comparison of correlations found between normal school records and success in teaching and between intelligence test scores and teaching in the two investigations could not be justified. Obviously, selection was no small factor in causing a group of normal school graduates to gravitate to a town or small city system. This one fact alone would tend to suggest the explanation for the absence of any close relation, the lack of appreciable correlation between them.

Meriam[2] found a correlation of + .443 between "practice teaching" and responsible teaching later, using different means of measuring the two performances. Had the same "measuring rod" been employed for both, as was done in the present study, the probabilities are that the relationship would have been somewhat closer. Moreover, had all the "practice teaching" been done in the same school of superior grade, in which actual school conditions are more closely approximated, chance errors would have been further reduced with a corresponding increase in the correlation.

Kelley's[3] study of the progress of a group of first-year high school pupils in relation to the elementary school record, teachers' estimates, and their reactions to special tests, although in a different field or at a lower stage, is the only major investigation, so far as the writer is aware, in which the regression equation has been previously applied on a large scale and with somewhat analogous results. To this study, reference has been made from time to time in earlier sections.

In the absence of a scale of performance in unit values for all ages and for all stages of development, the one best and safest available method of treatment is that which considers them from the standpoint of the group and which makes the mean performance the reference point for each individual score or measure. Normal ability is then the ability represented by the average for the group as a whole, and each individual is considered as so-much-above or so-much-below that point. Later comparison becomes a matter of measuring the individual's ability or performance by the standard (or what is judged to be the standard) for his or her particular group,—the person in every case being compared with the population of his or her own kind and of his or her own stage of development.

---

[1] Knight, F. B. *Qualities Related to Success in Teaching.*

[2] Meriam, J. L. *Normal School Education and Efficiency in Teaching.*

[3] Kelley, T. L. *Educational Guidance.*

In the absence of a single test or set of measures adequate to guidance needs, effectual combinations become increasingly necessary. Different tests and different (often dissimilar) measures are unequally related to the end-result,—the various types of achievement. Mere summation of unlike facts, unequally related, would not satisfy the conditions of logic or of science. And statistical determination is a modern necessity,—for only by mathematical treatment of related facts, each considered in connection with every other one and in relation to the end-result, can its real relationship and significance be known and its meaning grasped. Only by virtue of such independent relations does the one have bearing on the other, and in proportion to the amount or degree of this relationship should it have weight in the combination. Its contribution to the prediction is on the basis of, and in proportion to, the added increment it yields in the tested situation,—the value it adds to the prediction, experimentally determined. The value shown in one case, however, for a particular purpose, may be widely different from what may be found to obtain under other conditions and for a different purpose.

THE RESULTS

Pursuant to this method of investigation and treatment of facts discovered, the following conclusions may be safely drawn:

1. *Intelligence* as measured by mental tests reveals a relation to achievement in school and success in teaching as signified by a correlation of approximately +.54 and +.43, respectively. After allowing for all other measured influencing factors, a residual correlation (though small) still remains between test ability and

accomplishment. There appears to be a point below which good intentions and other traits, however worthy and commendable, do not compensate for this lack of a modicum of intelligence in excess of that possessed by a few high school "graduates."

2. *Personality* as measured by estimates of teachers exhibits a relatively close relationship to an individual's success as student and to her achievement as teacher, to the extent indicated by correlations of +.72 and +.61; and those traits considered important to the teacher's equipment are present in the recruit and identifiable to the degree and amount suggested by the latter coefficient.

3. *The first half year's record* portends about equally well with personality the end-result. Those activities which make for success during the early weeks and months make for similar success throughout the two years. "Exceptions" are found, but relatively few in number; this fact may be due either to a slower type of adaptability with uncertain adjustments to new and unfamiliar situations, or to an exploratory attitude leading to over-participation in circumambient activities with consequent inattention to the central requirements of the teacher training institution.

4. While all three sets of measures revealed appreciable relationship to the different types of achievement, no one alone yielded a sufficiently high correlation with any one type of success to serve as an adequate basis for prediction. For the purpose of predicting an individual's achievement, either in school or in the classroom outside, a more comprehensive basis than any single factor investigated in the present study is required. The more factors included, the closer the prediction to the actual outcome. Test-ability and Personality combined in a regression equation appear to promise a fairly adequate basis for forecasting a student's level of achievement; and these similarly combined with the complete school record serve approximately equally well for the teacher's level of success,—to the degree represented by correlations of +.746 and +.748, respectively.

5. Preparatory school records apparently possess but little less value for guidance purposes in the normal school than that found for the elementary school record in predicting success in high school work.[1] Such data, to be of greatest value, should be recorded

---

[1] Kelley, T. L. *Educational Guidance*, p. 84.

systematically and preserved in great detail. To insure practical uniformity, the Department of Public Instruction could profitably inaugurate a system for filing the complete annual record of every high school student, not only the "grade" or level attained in each subject pursued, but also the amount of the more important *personality-traits* exhibited, the latter estimated by the teaching staff. The individual-to-individual comparison is probably the best technique available for judging the amount of the quality present in the pupil. A transcript of the whole should be supplied as an essential part of the credentials to the higher institution to which application for entrance is made.

6. The combination of Tests and Personality might well serve to determine the fitness of candidates for admission to teacher-training institutions, as well as to colleges and universities,—not alone, possibly, but as supplementary to the preparatory school record, which is itself probably the most significant data, when available, concerning the prospective student, because it gives a measure of his or her ability as reflected in a wide array of special facts. Admission might well be denied to individuals who fall below –2 S.D. in a typical freshman group, without appreciable injustice to them or prejudice to the profession. In the absence of comparable high school records, elimination could be safely practiced at the end of the first half year on the basis of Test-ability, Personality, and First-Semester-Marks, using the same lower limits as above. Cases between –2 S.D. and –1.5 S.D. should provoke further and fuller investigation,—further testing and inquiry into the social background, individual purposes, bents, etc., —before final guidance is attempted.

The position or place of the individual in these measures or in combinations of them will serve as a criterion for judging later progress. Whenever a wide divergence between ability and accomplishment is discovered, the latter failing to keep reasonable pace with the former, attention may be called to the discrepancy and explanations sought with improvement expected, unless mitigating circumstances for failure are shown.

FURTHER STUDY

While the present investigation reveals relationships of measures and outcomes of a relatively high order, the results may not be considered above the plane of suggestions until confirmed by a more thoroughgoing study of the whole problem. The method followed is the correct one for the purpose, in the main, but the data available were not adequate to, nor would they justify, sweeping conclusions or wholesale recommendations. These must await a more far-reaching survey—the coöperative project of a corps of investigators, in several institutions, and of many teacher-recruits and educational practitioners in the field. Three lines of inquiry which seem to offer much, those richest in possibilities, for guidance and direction of the future teacher, may be thus indicated:

1. A larger number of traits and qualities considered important to the teacher's personality might profitably be studied by the methods pursued in the present investigation, with the added judgments of the high-school teaching staff and estimates of fellow-students. Such an inquiry should be conducted in a number of teachers' colleges and normal schools, varying in size of student body and faculty and practice teaching facilities, and located in different parts of the country.

2. The writer is not unaware of certain limitations inherent in the present method of measuring human traits,—the difficulty of expressing an individual's possessions in intelligible units of amount, the differences in experience and ability of "judges," and the multiplication of estimates necessary to insure reasonable validity for the judgments made. For these and other reasons, research in the field of objective measurements, although inconclusive[1] at present with the moral and emotional phases of life activities, is by no means barren in possibilities, and may, in fact, prove to be a most fruitful field for both diagnostic and prognostic purposes.

A person's reactions to ingeniously devised situations may be a truer measure of his or her humor than the judgments by acquaintances. Concrete situations being exceedingly difficult to arrange and control, and generally impracticable on a large scale, "paper" performances may prove a fit substitute,—the language accompaniment to the "sense of humor" may be a really reliable index of its presence and strength. Reactions to unique completion sentences, preferences for "graded" selections of prose and verse, supplying paired associates, choice of reasons to *satisfy* a certain state of affairs, selection of a word, among several, to *satisfy* the statement, seem suggestive of possibilities here; and it may be that other traits will lend themselves to objective treatment as well.

---

[1] Downey Individual Will-Temperament Test. World Book Company.

3. To measure a teacher's worth by the impression and judgment of the supervisory staff may be far different from what the facts in the case, were they known, would warrant. If it may be assumed that the value of a teacher is in proportion to the desirable changes wrought in the attitudes and activities of those for whom and with whom she works,—the growth and development of the pupils, the shared relations of co-workers, and enlarged point of view of patron and public,—such a measure may be, and frequently is, but a partial picture of her real worth. This judgment may be accepted as a fairly precise index of the effect of a teacher's influence on her fellow-teachers and a partial index of the increased enrichment of the community life due to her presence; but it may not follow that proportional changes in pupils are concomitant results. The correlation may be positive and appreciable, but a safe guess would place it well below a 100 per cent perfect (or plus 1.00).

Measurements, by means of objective tests, of the pupil's progress, which may justly be accredited to the teacher's influence, seem necessary to a just evaluation of his or her worth. The better known and more widely used tests and scales on the common branches, together with tests on ideals and civic life,[1] rightly applied and in sufficiently large numbers, should serve to supplement the judgment of principal and supervisor, and thus give a more nearly correct value than is possible on the basis of the latter alone. This is a line of inquiry worthy of pursuit.

Confidence is abundantly justified, in the progress made thus far, but many comprehensive investigations may be required before the major problems of educational and vocational guidance have been solved; but only through experimentation with life situations, as we find them, and with facts as they are, will truth stand revealed, uncontested and incontestable.

---

[1] Upton, S. M. and Chassell, C. M. "A Scale for Measuring the Importance of Habits of Good Citizenship," *Teachers College Record,* January, 1919.

Psych Somers, Grover Thomas
S694 Pedagogical prognosis

PLEASE DO NOT REMOVE
CARDS OR SLIPS FROM THIS POCKET

UNIVERSITY OF TORONTO LIBRARY